

Crick's immodest ambitions

As scientists and others worldwide celebrate the 50th anniversary of a profound discovery, one of the discoverers persists with another challenging quest that, while still a work in progress, also deserves celebration.

Winning a Nobel Prize is a life-changing event. It instantly accords influence and status to its recipient in domains far outside the realm of the discipline being honoured. Some use this status to influence science policy. Many use it to attract donations or investment to their institutions or other worthy causes. More rarely, a laureate will apply the same principles that guided them to Nobel honours to continue their groundbreaking career in another scientific field. Francis Crick is one of the latter: decades after his discoveries about the root of life, DNA, he continues to investigate deep scientific questions by turning his talents to the root of consciousness, the brain.

Nature has not only had the privilege of publishing Crick's seminal contributions to our understanding of the molecular basis of life, but for the past two decades has also had the pleasure of publishing papers from him on various topics in cognitive neuroscience. Often in collaboration with leading neuroscientists, he has focused on the functional organization of the brain. For example, he has tackled the question of why we sleep, by considering the brain as a neural network whose organization may need to be periodically refreshed. He has described the impoverished status of studies in human neuroanatomy. And with Christof Koch he has written extensively about visual awareness and consciousness. In particular, Crick has championed the study of the neural correlates of consciousness: to use an example from his book *The Astonishing Hypothesis* (Simon & Schuster, 1994), you perceive something to be red "if and only if certain neurons and/or molecules in your head behave in a certain way".

When Crick turned his attention to neuroscience, neurobiological investigations into consciousness were far from mainstream. Philosophers regarded consciousness as their exclusive domain, only occasionally ceding territory to neuropsychologists, who brought insights by studying patients such as those with blind-sight or split brains. Fundamental neuroscience was not seen as having much to offer, even by its own practitioners. The great nineteenth-century neurophysiologist Emil Du Bois-Reymond famously declared the scientific exploration of consciousness impossible ("*ignoramus et ignorabimus*"), and that view persists among some neuroscientists even today.

Making connections

But Crick has helped to bring consciousness and neuroscience together, emboldened by the confidence that led James Watson to observe mischievously that he had "never seen Francis Crick in a modest mood". Many great thinkers have turned to the human mind as the last great frontier of scientific endeavour. But few have done it with the rigorous attention to the neuroscience — from single-unit physiology to cognitive psychology — that must inform any biologically based theory of consciousness.

Having catalysed one scientific revolution by answering the correctly posed reductionist question, Crick approached consciousness through a similar search for the appropriate questions to ask. Rather than endlessly debating a precise definition of consciousness, Crick and his colleagues have focused on visual awareness as a phenomenon that is both central to the human experience and

amenable to research with animal models. Rather than turn to vague ideas of 'emergent' brain properties, they have taken a first step to specificity, proposing that the activity of specific populations of neurons correlates with specific conscious phenomena.

Neuroscience still lacks the tools to accumulate data rapidly, so it is too early to assess the impact of Crick's specific proposals. Nonetheless, his questions are concrete enough to be taken as a challenge rather than a discouragement. It is not enough to know how a certain percentage of the neurons in a particular area respond during visual perception; we must identify the neurons involved, and know where they project and what neurons connect to them, before we can prove or disprove the idea that visual awareness is a function of higher areas of the cerebral cortex. Neuroscientists are constantly pushing the limits of the techniques available to them, of course, and Crick was not the first to point out that information of this kind is needed. But his powerful advocacy of the neural-correlates approach has helped to stimulate the community to develop new methods of answering the questions that he posed and of asking still more sophisticated ones.

All in the mind

More generally, neuroscience suffers from a shortage of strong theories. We would all agree that cracking the mysteries of the brain requires new hypotheses, but theorists are often undervalued. There seems to be simultaneously too much data to manage and yet too little data that can rigorously constrain a particular theory. It is too easy, many feel, to come up with a plausible neural explanation for a particular cognitive phenomenon, but too difficult to support or falsify it definitively. The framework for studying consciousness that Crick and his colleagues continue to articulate is one example of a guiding theoretical force that, once appropriate experimental tools become available, will not remain an unfalsifiable 'just so' story.

The insatiable public appetite for explanations of the mind is too often met with the intellectual equivalent of junk food. In contrast, Crick has brought the neuroscience of consciousness into the realm of the educated general public with *The Astonishing Hypothesis*. The challenge of bringing a unified theory of mind and brain to a population that largely believes the two to be separate entities requires both prominence and a clear voice.

The ideas on the neurobiological basis of consciousness that Crick has put forward over the years are not tightly bound into one consistent theory. They are, instead, explorations in a vast ocean of possibilities. They encourage neuroscientists to meet the challenge of this daunting topic. Like the question 'What is life?' — posed by an earlier Nobel laureate, Erwin Schrödinger, who was also unafraid to step outside the discipline in which his name was made — 'What is consciousness?' may be recast as more information is accumulated. If so, it will have made the transition from philosophy to science.

Crick carries forward the courage of his conviction that consciousness is worthy of serious scientific investigation by being both an insightful critic and an advocate of neuroscience. In doing so, he gives us grounds to hope, and indeed to expect, that the problem of consciousness is best placed within the scientific domain. ■





War in Iraq
Aid agencies express
fears over conditions
in Basra
p459



Family trees
Geneticists join the
hunt for human
ancestors
p460



In the hot house
University faces legal
action over funding
for Biosphere 2
p461



Under review
Vaccination deaths
force revision of
smallpox policy
p462

Researchers rage at tightened restrictions on US immigration

Geoff Brumfiel, Washington

Scientific leaders are increasingly fearful that tighter immigration procedures, introduced in the wake of the terrorist attacks on 11 September 2001, are threatening the United States' position as a magnet for the world's scientific talent.

Researchers from countries as diverse as Indonesia and Germany are now subject to detailed security checks and rigorous interviews. The clampdown covers first-time visitors to the United States and those returning to lab positions there — delaying trips by weeks or months, and deterring some from coming at all.

The consequences of the change, which intensified with the introduction last August of new visa guidelines for consular officials, could be far-reaching. "We are in a rapid transition, whereby the United States will cease to be the destination of choice for researchers," predicts Irving Lerch, director of international affairs at the American Physical Society.

Some researchers and officials outside the United States — such as those at top European universities — acknowledge that they could benefit from a protracted reduction in the flow of scientists into US institutions. But publicly, at least, they draw little comfort from the situation.

"I'd prefer a world in which individuals make free decisions about where to go and



Locked out: Chinese students protest over the US embassy's refusal to grant them a study visa.

work," says Robert May, president of Britain's Royal Society. "We need to keep Britain and other European destinations attractive for scientists — but not as second-choice countries."

Last week in Congress, the House Science Committee held hearings to address scientists' concerns. "The current situation is untenable," argued the committee's chairman, Sherwood Boehlert (Republican, New York). "Foreign students fill our graduate

programmes; foreign scholars fill our faculty and laboratory positions. These people are a vital source of new ideas and perspectives." But not all committee members agreed, with some voicing satisfaction at the exclusion of foreign scientists (see below).

The overall scale of the shift is difficult to quantify, but some indicators suggest that it is significant. At the Human Frontier Science Program, which funds international collaborations between biologists, for example, the

Congressmen unmoved by foreigners' plight

Worries about the entry delays being experienced by foreign researchers coming to the United States were the avowed topic of a hearing of the House Science Committee on 26 March. But most of the members who turned up had different concerns on their minds. They wanted to know what is being done to make the United States more secure — and to lessen its dependency on foreign scientific talent.

"I am not particularly concerned that there is a little bit more red tape," Phil Gingrey (Republican, Georgia) told the scientific leaders who testified at the hearing. "Our security is more important than your convenience."

Ralph Hall (Democrat, Texas), the senior

minority member on the committee, said that America's dependence on foreign students bothered him almost as much as the country's dependence on foreign oil.

And Dana Rohrabacher (Republican, California) suggested that the domestic shortage could be remedied if US students were given preference over foreigners for places in US graduate education programmes. "There are scientists from communist China swarming all over Los Alamos lab," he said, "and when the Chinese start building rockets efficiently enough to hit any American city, we can start blaming this open exchange that we've had between scientists and our universities."

Some research leaders at the hearing were discouraged by the lack of sympathy expressed by committee members towards the plight of foreign scientists. But Wendy White, who directs the National Academies' Board on International Scientific Organizations, was pleased just to see the issue raised. "I was heartened by the fact that there were hearings at all," she says.

Hall and Sherwood Boehlert (Republican, New York), the chairman of the committee, said at the hearing that they have written to the General Accounting Office requesting a study to assess the backlog and to review visa security-check procedures.

Geoff Brumfiel

percentage of fellowship applicants who want to work in the United States has fallen from 75% in 2001 to 55% this year.

"The arbitrariness of the US immigration machinery has increased to a disturbing degree," laments Erwin Neher, director of the Max Planck Institute of Biophysical Chemistry in Göttingen, Germany, who won the 1991 Nobel Prize in Medicine for his work on ion channels in cells. Several foreign scientists at the institute — from countries such as South Korea, India and Hungary — and at the nearby European Neuroscience Institute have been refused US visas altogether, Neher claims. "It is awful," he says. "Our PhD student exchange programme with Stanford University is suffering badly."

US research programmes, such as that of the Center for Nonproliferation Studies in Monterey Bay, California, are also feeling the pinch. Several researchers from the countries that the centre needs to work with, including China and parts of the former Soviet Union, have been denied entry to the United States. "I'm supportive of greater controls, but they are using a very blunt instrument and they are weeding out the wrong people," argues Clay Moltz, a project director at the centre. Before 11 September 2001, Moltz says, he could usually talk to the US embassy in the country concerned to move an application along. Now the embassies don't return his calls. "There is no communication at all. You file these things and wait," he says.

Also affected are public-health programmes, particularly those aimed at tackling infectious diseases such as HIV. Chris Beyrer, who directs the AIDS International Training and Research Program at Johns Hopkins University in Baltimore, Maryland, says that visa delays have occurred in all of the dozen or so countries with which his programme deals. Practically every applicant from China, where HIV is rapidly taking hold, has encountered long delays, he says. US embassies in Uganda and Thailand are entirely rejecting visa applications from about 1 in 10 of those seeking training at Johns Hopkins, Beyrer says. "We need to train scientists in Africa, and some of them have to get to the States," he says. "Antiviral therapy for AIDS is not something you can learn from a handout."

Visitors to the United States have faced tighter entry controls ever since the 2001 terrorist attacks, but the problems intensified last August, when the Department of State broadened its guidelines covering visiting researchers from "sensitive countries" to encompass those from all destinations. Under the guidelines, candidates for work and study visas whose applications mention any of a wide range of topics — including chemical engineering, biochemistry and microbiology — face possible interviews by consular officials (see above right).

After these interviews, officials can elect to

Physicist discouraged by heightened security

When Wolfgang Caliebe tried to return to work at the Brookhaven National Laboratory in New York state after a Christmas break at home in Germany, he was surprised to find himself subjected to a security review.

In previous years, Caliebe — a native of Kiel who has spent ten years as a student and researcher in the United States — has never had to wait more than a week or so for his visa to be renewed. But this time, when he applied for its extension, he was told he would have to stay in Germany for at least a month and be interviewed by an officer at the US consulate in Frankfurt. He eventually got back to Brookhaven, but says that he has now given up on a career in the United States and is searching for a job back home.

Caliebe is one of a handful of physicists who provide technical support at the National Synchrotron Light Source, a facility used by some 2,500 scientists to image molecular structures and materials. "If one of us is gone, we are very thinly spread," he says. Furthermore, Caliebe is the only one dedicated to caring for an elastic-scattering instrument at the source. "If they have a problem and I'm not around, the instrument just sits around and collects dust," he adds.

With the help of Brookhaven officials, Caliebe was able to lobby for an earlier date for his interview that would delay his return by just two weeks. But in the days before the meeting, he was very much on edge. "I didn't know why they had invited me for an interview," he says. "Sometimes, I couldn't sleep at night."

When the time came, Caliebe was confronted by an officer who asked a series of questions about his work. "I described how the storage ring works, but she didn't understand that," he says. He says that the officer's attention only picked up when he mentioned radiation. "I was scared at that point," he recalls. "I felt worse than I did during my PhD defence."

In the end, Caliebe got his visa, but he says that colleagues at Brookhaven have fared less well. One student from Indonesia had to wait three months to return to the United States. "Now he has to go regularly to the police station and tell



Wolfgang Caliebe: questioned before his return.

them, 'Yes, I still work at Brookhaven, and I'm a nice guy'," Caliebe says.

Caliebe's experience has persuaded at least one German graduate student, who sought advice from him on postdoctoral openings, to steer clear of the United States. "I described to him the situation I was just in, and he said: 'OK then, I'd better look in Europe'," Caliebe says.

Geoff Brumfiel

send the application to Washington DC for an interagency security review. Hundreds of scientific applications have been referred for such reviews, partly because consular officials lack scientific training. And with several agencies — including the Federal Bureau of Investigation, Central Intelligence Agency, Department of Homeland Security and Department of Defense — having to check each case, this process can take months.

"I'm aware of the problems, and I think they are serious," says John Marburger, President George W. Bush's science adviser. Marburger says that his staff at the White House Office of Science and Technology Policy is meeting "almost every day" with

agencies throughout the government to speed up the visa process. Marburger hopes to streamline the reviews, and to involve more scientifically qualified people in them. "I think we can do a better job without increasing the risk of terrorism," he says.

But most observers don't expect the situation to improve any time soon. The United States' top priority is national security, points out one official in the Department of Energy, who works with visa applications. "No one wants someone to point a finger at them and say 'You just caused the death of 3,000 Americans,'" the official says.

Additional reporting by Quirin Schiermeier in Munich and Jonathan Knight in San Francisco.

Aid workers fear impending disaster in Basra

Quirin Schiermeier and Tobias Kramer

The siege of Basra seems to have caught out health officials working with international aid organizations. As *Nature* went to press, coalition forces had surrounded Iraq's second-largest city, and water supplies to around half of the 1.7-million population had been cut off for over a week. But aid agencies expected humanitarian problems to be concentrated in northern Iraq, and so lack the staff and equipment to contain any problems in Basra.

Iraqi officials say that there were no reports of cholera, typhoid or dysentery in the city last week. But experts with the World Health Organization (WHO) warn that deteriorating sanitation, water contamination, and increasing heat and humidity mean that there is an extremely high risk of an outbreak of diarrhoeal diseases. The hundreds of thousands of children in the city, many of whom are suffering from malnutrition, are particularly at risk.

"We are on the highest alert," says Ghulam Popal, head of the WHO's office in Baghdad, which coordinates the activities of the organization's 327 Iraqi medical staff.

The WHO has 15 emergency health kits — each containing basic needs for 10,000 people for three months — in place in Baghdad, and others are stored in northern Iraq. But the organization has no staff or kits in Basra, so the population there will have little or no access to the equipment, which includes antibiotics, dehydration solutions



The Red Cross is currently the only aid organization with access to the beleaguered city of Basra.

and water-purification tablets — all of which are vital to combat diarrhoeal diseases.

"You go where you think you'll be needed, and where it is reasonably safe for your own people," says Barry Sandland, a Brussels-based spokesman for Médecins Sans Frontières, a charity that has sent a team of six doctors to Baghdad. "But at the end it is just guesswork."

WHO officials believe that few people in Basra have access to public-health services. With the city's hospitals overloaded with wounded civilians, possible epidemics are likely to be a secondary consideration for the local health authorities, says Jim Tulloch, the WHO's regional health coordinator in Amman, Jordan.

A WHO rapid-response team in Baghdad is waiting for permits from the Iraqi authorities to travel to Basra, some 500 kilometres south of the Iraqi capital. The group will then have to decide if it is safe to make the journey. "We need immediate access to the city to assess the risk of an epidemic and, in the case of an outbreak, identify its cause and start medical treatment," says Popal.

Currently, the only international aid organization with access to Basra is the Red Cross, thanks to safety guarantees from the Iraqi army and the coalition forces. According to a Geneva-based spokesman, attempts to repair Basra's main water station are under way, but the water supply was still worryingly poor at the beginning of this week. ■

China joins investigation of mystery pneumonia

David Cyranoski, Tokyo

The World Health Organization (WHO) has confirmed that it has managed to enlist China in the international investigation of the mystery pneumonia known as severe acute respiratory syndrome (SARS).

But even as agreement was reached, leading Chinese researchers clashed with WHO investigators over the probable cause of the illness.

SARS arose in East Asia and, according to WHO figures on 31 March, it has so far spread to 13 countries, infecting 1,622 people, 58 of whom have died. One of the recent casualties was Carlo Urbani, an Italian physician working for the WHO, who identified the first case of the disease while treating patients in Vietnam.

Researchers and public-health officials now assume that SARS is the same disease as that which struck the Guangdong province in southern China late last year. Initially, Chinese officials said that the epidemic had

infected about 300 people and had petered out in February (see *Nature* 422, 247; 2003). But last week, they admitted that by the end of February it had infected at least 806 people, causing 34 deaths.

An international investigation coordinated by the WHO has identified two virus families — coronaviruses and paramyxoviruses — as the possible cause of the epidemic, although some believe that the illness may involve a combination of both families. But Tao Hung, a researcher at the Institute of Virology at the Chinese Academy of Preventative Medicine in Beijing, who is heading the Chinese investigation of the epidemic, says that his results point in another direction — a *Chlamydia*-like bacterium. He says he has found traces of *Chlamydia* in samples from organs taken from five people who died from SARS and from 30 infected patients.

But researchers inside and outside China dispute that the normally non-lethal

Chlamydia could exact such a heavy toll.

"It would go against common sense," says Masato Tashiro of the National Institute of Infectious Diseases in Tokyo, one of the WHO's investigators.

"From a clinical point of view, it cannot be *Chlamydia*," says Nan Shan Zhong, a doctor at Guangzhou Medical College in Guangdong, who is advising the government on the outbreak in the province. He adds that antibiotics against *Chlamydia* have proven ineffective in treating the syndrome. "I do not believe the disease is really under control," he says.

In the agreement reached on 30 March, a WHO delegation led by John Mackenzie, a virologist at the University of Queensland, Australia, convinced China to become a full participant in the WHO operation. This will mean full updates, sharing resources, and allowing WHO officials to visit doctors in Guangdong — something that has been repeatedly promised but has yet to occur. ■

Genetics to unlock secrets of our African past

Michael Cherry, Stellenbosch

Archaeologists and geneticists could soon be working together to investigate the genetic diversity of our human ancestors in Africa, a meeting on the "Human Genome and Africa" was told last month.

Chris Stringer of London's Natural History Museum told the meeting, organized by South Africa's Human Sciences Research Council (HSRC), that the identification of well-preserved fossil hominids and analysis of their DNA could answer key questions about the origin of our species. It is widely accepted that humans evolved from hominids that left Africa during the past 200,000 years — but little is known about the location or genetic diversity of the African groups.

African fossil hominids have very diverse morphology compared with those from Europe, where only two distinct groups — the Neanderthals and Cro-Magnons (the progenitors of modern humans) — have been found. "The sheer size of Africa may mean that it could conserve diversity in genetic material, and innovations in human behaviour, better than other areas," Stringer says.

In Europe, Neanderthal DNA has already been sequenced successfully, but ancient DNA has not yet been extracted from African hominid fossils. According to Julia Lee-Thorp, an archaeologist at the University of Cape Town, very little organic material is

preserved in African fossils more than 10,000 years old, because preservation requires cool, dry conditions. "One would have to look specifically for sites with extraordinary preservation conditions," she says.

Archaeologists have some potential sites in mind in South Africa. "I think that there would be considerable interest in funding this kind of research," says meeting organizer Wilmot James, director of the HSRC's social cohesion and integration programme.

But permission for excavations, previously the responsibility of the National Monuments Council, was recently devolved to the country's nine provinces. Most provinces lack the infrastructure to administer this mandate, resulting in an effective moratorium on excavations. There is also currently a moratorium on the export of fossil samples for research, requiring the genetic analysis of any samples to be done locally.

Stringer thinks the great diversity of relatively recent hominids in Africa reflects genetic bottlenecks generated as a result of the climate change that occurred between 400,000 and 10,000 years ago, forcing populations to withdraw to small refuges. The surviving populations in these areas could have evolved rapidly in response to sexual selection based on, for example, facial features — leading to a diverse-looking collection of hominids across the continent. "It is quite likely that these ancestors did



DNA from artefacts such as this 100,000-year-old jawbone could shed light on hominid diversity.

R. YATES

not closely resemble any extant racial group," says Stringer, "as modern regional features are not seen in skulls more than 30,000 years old, from either Africa or other continents."

Several speakers at the conference, which was held on 19–22 March and attracted 350 scientists and officials from 17 African and 13 other countries, called for Africa to move faster to use genomics in medicine and social science. Gordon Dougan, director of the Centre for Molecular Microbiology and Infection at Imperial College, London, said that African scientists need to master genomics so they can develop their own vaccines and drugs, outside the control of multinational companies. ■

Axeing of website article sparks row at Max Planck

Alison Abbott, Munich

The Max Planck Institute for Plant Breeding Research in Cologne has removed the detailed description of 'intelligent design' from its website, following complaints from scientists that it was inconsistent with the laboratory's scientific mission.

The article, which was posted by Wolf-Ekkehard Lönnig, a theorist at the institute, discusses the idea that an intelligent force must be responsible for the origin of the

Universe and for the diversity of life forms. Known as intelligent design, this theory rejects natural selection, and has been portrayed by its opponents as a 'front' for creationism (see *Nature* 416, 250; 2002).

Earlier this month, Peter Gruss, president of the Max Planck Society, asked the four directors of the Cologne institute to provide a scientific justification for Lönnig's pages. Lönnig posted the material five years ago, and the site has since received over 35,000 hits. A disclaimer identifying the article as a personal opinion was added in 2001, following earlier complaints.

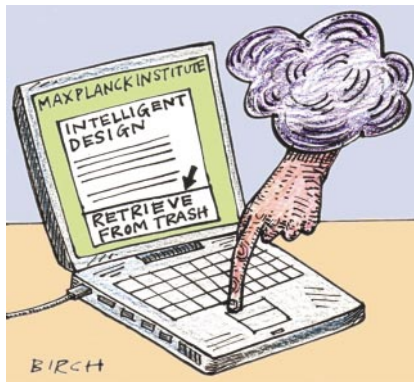
"Only scientific issues should be discussed on a Max Planck site," says Gruss. And last week, Lönnig's pages were removed from the institute's site, pending a directors' meeting on 28 April to determine their fate.

Ulrich Kutschera, an evolutionary biologist at the University of Kassel, has campaigned against the presence of the material on an official Max Planck website, branding it "pseudoscience". "It is fine as a personal opinion expressed on a personal website, but not on the official site of a

scientific organization of international status," he says.

Many evolutionary biologists share Kutschera's concerns: Axel Meyer of the University of Konstanz, for example, says that he was "shocked" by the contents of the pages. But others, such as Diethard Tautz at the University of Cologne and Svante Pääbo at the Max Planck Institute for Evolutionary Anthropology in Leipzig, are more circumspect, saying that independent opinions should be permitted. Tautz, however, says it might be more appropriate for such opinions to be aired at "an institute of philosophy" than at the Max Planck.

Lönnig is displeased by the removal of his discussion. "No one is happy when someone switches off the information flow of what he thinks is right," he says. And Heinz Saedler, one of the institute's directors, who has supported Lönnig and published jointly with him, says that although he doesn't believe in intelligent design himself, he enjoys discussing it with Lönnig. ■



Report fuels criticism of UK research council

Natasha McDowell, London

For years, critics of Britain's Medical Research Council (MRC) have quietly grumbled that the prestigious funding agency doesn't plan its priorities carefully enough, or communicate properly with the community of researchers that it supports. Now the detractors are claiming vindication after a scathing report from a parliamentary committee endorsed their views.

The report, released on 25 March, says that the MRC has grown distant from many medical researchers. The committee accuses the council of "inconsistent and inadequate communication"; and of allocating too much money to big projects, such as the UK Biobank genetic database, leaving itself unable to fund as many individual grants as it had led researchers to expect (see *Nature* **418**, 714; 2002). "The recent success rate for the MRC's grant applications has fallen to levels that are unacceptable," the report says, adding that researchers' anger at this is "entirely justified".

George Radda, a biochemist and chief executive of the MRC, says that many of the charges levelled by the House of Commons' Science and Technology Select Committee result from "misunderstandings" about what the agency is doing.

But several medical researchers contacted by *Nature* say that the charges are broadly justified. "In my view — and that of most of my colleagues — the criticisms are just about on the mark," says David Colquhoun, a pharmacologist at University College London.

"A lot of what is in the report is valid and fair, and the MRC needs to take note," says David Price, a physiologist at the University of Edinburgh and one of the scientists who testified before the committee.

"The MRC has a good track record, but in the past few years it has gone to pieces," says Ian Gibson (Labour, Norwich North), the committee's chairman. The harsh language in the report reflects the views of organizations and senior scientists who testified before the committee, he says. He adds that, in his view, the MRC is run by a small, London-based group that is not fully representative of medical researchers around the country.

Radda defends the decision to back the UK Biobank, to which the MRC has pledged £20 million (US\$32 million) over seven years. "We have a responsibility to set up national facilities that support science in the longer term," he says. But he says that the MRC should have done more to inform scientists of its financial situation. "We now have a very detailed statement on our website about funding between 2003 and 2006," he adds.

And William Stewart, a microbiologist and former chief scientific adviser to the

British government, says that Gibson's committee has in the past been overly critical and might have been more constructive in this case. "They should be fighting for more money for science," he comments, "but there's no mention here of the MRC's need for more funds."

The government is due to respond to the report within two months, and its officials had no comment last week on what it will do about the barrage of criticism levelled at the council. But Peter Cotgreave, director of the pressure group Save British Science, thinks the findings will encourage more energetic communication from the MRC and discourage further commitments to large projects. "If they commit money long-term next time, having had these problems pointed out, they will be asking for trouble, and may not get funded," he suggests. "The other research councils will argue that they should get more money because they know how to spend it." ■



George Radda, head of the MRC, defends the allocation of money for longer-term research.

Biosphere owner sues sponsor

Virginia Gewin, Portland

The owner of Biosphere 2 — an experiment to build a carefully controlled ecosystem in a massive Arizona greenhouse — is suing Columbia University in New York over its plan to withdraw support from the facility.

Decision Investments Corporation, a company controlled by the Texan billionaire Ed Bass, who built Biosphere 2, filed a lawsuit against the university in Arizona's superior court on 21 March. Two months ago, the company was notified that Columbia wanted to get out of its ten-year, \$20-million contract to manage the facility until 2010, by June (see *Nature* **421**, 466; 2003).

The conflict arises as Biosphere 2 takes steps towards attaining the scientific credibility it has sought since its construction 15 years ago. Earlier this year, for example, chemists there demonstrated that elevated atmospheric carbon dioxide levels could offset negative air-quality effects associated with planned forests (T. N. Rosenstiel *et al.* *Nature* **421**, 256–259; 2003).



Biosphere 2 is now the subject of a court battle between its owner and Columbia University.

The lawsuit alleges that Columbia is in breach of contract by abandoning plans for new education programmes, an extra lab and the recruitment of half-a-dozen senior researchers at Biosphere 2. "For years, Columbia has affirmed that the education and science tracks were proceeding with much success, and then, out of the blue, they want to cease funding the project they designated as their 'western campus' three years ago," says Martin Bowen, vice-president of Decision Investments. A spokeswoman for Columbia declined to comment on the lawsuit.

A statement by Decision Investments ascribed Columbia's move to a recent change of leadership — economist Jeffrey Sachs took over last April as director of Columbia University's Earth Institute. But Sachs says the institute will maintain its commitment to ecological research, and that any speculation to the contrary is "totally unfounded and completely misinformed".

Biosphere 2 managers say they plan to keep operating the facility and to find alternative sources for its annual operating costs of about \$1.5 million. They say that European researchers, including groups from Jena, Germany, and Edinburgh and remain committed to the experiment.

Joe Berry, a Biosphere 2 ecologist based at the Stanford University branch of the Carnegie Institution of Washington, says he thinks that the facility could one day sustain itself through grants if a critical mass of good-quality scientific output is reached. ■

Fresh funding hike mooted for US biomedical research

Washington Fears that the recent rapid growth in US biomedical spending has reached its end could prove to be premature. A vote on 26 March saw senators endorse the idea of increasing next year's funding for the National Institutes of Health (NIH) to \$29 billion, an 8.4% rise on its 2003 budget.

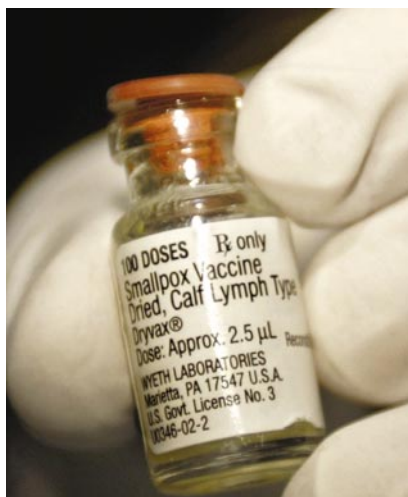
Research advocates have feared that the NIH may get short shrift next year, because President George W. Bush's February budget announcement included only a 2% increase (see *Nature* **421**, 565; 2003). Many advocates say that a bigger increase is needed to sustain the momentum generated by an effective doubling of the agency's funds since 1998.

The vote does not guarantee that the NIH will get its 8%. A final decision will come in autumn, when the Congress is scheduled to write an appropriations bill that will distribute money to all federal agencies.

Heart-attack deaths raise fears over smallpox jabs

Washington Three healthcare workers have died of heart attacks after receiving smallpox vaccinations, leading a government advisory committee to revise its recommendations on who should receive the shots.

The link between the deaths and the vaccine is unclear, but all three victims had a history of heart disease. The Advisory Committee of Immunization Practices said on 28 March that people with heart disease or those with three or more risk factors for it, such as diabetes, should not get the vaccination. Some 6% of healthcare workers could now be excluded. Healthcare professionals began to be vaccinated earlier this year after fears grew that smallpox could be used in a bioterror attack.



Smallpox vaccine is being given to US healthcare workers to protect them from bioterror attacks.

The advisory committee considered excluding anyone aged over 50 from being vaccinated, but feared that it would cause too much damage to the programme. Some doctors recommended that the entire programme be put on hold until the link between the vaccine and the deaths could be cleared up.

Canada drives medical protein-structure hunt

Toronto The three-dimensional structures of more than 350 human proteins are to be deciphered over the next three years by an Anglo-Canadian partnership.

Four leading Canadian funding agencies, together with British biomedical charity the Wellcome Trust and pharmaceutical company GlaxoSmithKline, will make the first results of the £40-million (US\$63-million) programme freely available at the end of this year. The work of the Structural Genomics Consortium will be carried out in laboratories at the

universities of Oxford and Toronto. They will focus on proteins of medical relevance, such as those that are associated with cancer, neurological disorders and malaria. Aled Edwards, a proteomics researcher at the University of Toronto and director of the consortium, says that the results will enable scientists "to put the genome to practical use".

The initiative marks the first time that the four Canadian funding agencies, which include Genome Canada and the Canadian Institutes of Health Research, have worked together.

Columbia data hint at wing damage during launch

Washington Data from a sensor package tucked inside the doomed space shuttle Columbia has given credence to the theory that the vehicle was damaged even before it began its fateful entry into the atmosphere.

The readings, recovered from a tape found during a ground search on 19 March, show that the shuttle's left wing began to overheat more than a minute earlier than previous sensor data had suggested. The heating was probably due to plasma gases entering the wing. The data support the idea that part of the wing's insulation had already come loose, probably when the wing was struck by pieces of foam insulation that fell from the shuttle's fuel tank shortly after launch.

Some NASA managers say they hope to launch a shuttle as early as this autumn. But Guy Fogleman, acting director of the agency's Bioastronautics Research Division, told the National Academy of Sciences' Space Studies Board last week that "the most optimistic date we've heard is early next year".

Cash crisis forces US genetics lab to close

Denver Tough economic times have forced the closure of the independent Eleanor Roosevelt Institute, a key player in the sequencing of human chromosome 21 (see *Nature* **405**, 311–319; 2000).

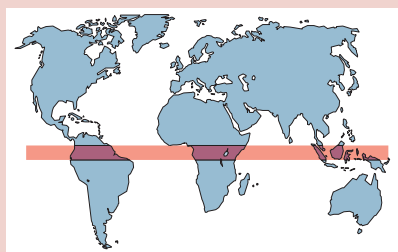
Operating costs for the institute, which is based in Denver, Colorado, and focuses on genetic research into conditions such as Down's syndrome, were around \$4 million in 2001. Charitable donations provided one-third of the total, with federal grants providing most of the rest. But the economic downturn has limited philanthropic income over the past year and a half (see *Nature* **419**, 765; 2002).

"It's really hard for a small institute to maintain itself in this economy," said David Patterson, the institute's president. Patterson adds that he has searched for an alliance with a university or other large organization, but that the current economic climate has scared away potential partners.

Mind your heads!

Rome Residents of equatorial countries will be keeping a nervous eye on the sky later this month. BeppoSAX, an Italian X-ray satellite, is scheduled to re-enter the atmosphere, and pieces of debris as heavy as 120 kilograms could rain down on regions below its path.

The 40 or so pieces of BeppoSAX that are expected to survive the re-entry should pose little danger to people on the ground, according to the Italian Space Agency (ASI), which estimates that atmospheric drag will cause the satellite to plummet to Earth on 30 April. ASI officials estimate the risk of somebody being hit at less than 1 in 5,000, but have warned people within the re-entry zone that the satellite's fuel and battery chemicals could be toxic. Indonesia



has the largest population in the zone — a band of more than 4° on either side of the Equator (red band on map above).

Ground controllers can't steer the satellite's re-entry, and won't even have a precise estimate of its final re-entry point until it begins to fall. On average, a large piece of space debris falls to Earth every 10 to 12 days, the agency points out.

How to clean a beach



As oil-spill specialists continue to tackle the *Prestige* slick, they are drawing on knowledge from decades of clean-up operations. John Whitfield reports from Spain's Galician coast.

From the coast road, the beaches of Lira seem as they should be: yellow sand and blue sea. But walk down to the tide's edge and things change. A whiff of petrol taints the sea spray. Water in rock pools has an oily sheen and boulders that should be wet and slippery have a tacky, tarry coating. After the oil-tanker *Prestige* spilt her cargo last November, these coves in Galicia, in northwest Spain, were a metre deep in a mixture of oil and sea water known to pollution specialists as 'chocolate mousse'. "There was no ocean, only oil," says Pablo Garcia, manager of the Stolt Sea Farm, an aquaculture company in Lira, the area that became known as Ground Zero of the spill.

The *Prestige* is the latest exhibit in the tanker hall of infamy. But while each new incident brings environmental destruction and financial loss, it also improves our understanding of how to deal with oil spills. This knowledge is hard-won — aggressive clean-ups have sometimes caused more damage than the oil. Government priorities can also clash with those of scientists. But such difficulties apart, a rough consensus on how to juggle the political, economic and ecological issues involved in clearing up oil

spills has begun to emerge.

Oil is much less damaging at sea than on shore, so the best option is to suck and skim a slick off the water using specially equipped ships, or break it up with chemical dispersants. Booms can also be used to protect the coastline. But the sea around the *Prestige* was too rough and much of the coast too exposed for booms to work, so there was little that could be done except watch the oil wash up.

The human touch

When oil arrives onshore, the question becomes how best to save affected plants and animals while minimizing damage to the surrounding ecosystem, and without running up a huge bill. In large spills, leaving nature to do the job is a bad idea. Even the oiliest shore will return to normal, but without human intervention this can take a long time. In 1974, the *Metula* spilled 50,000 tonnes of oil into the Strait of Magellan at the southern tip of South America. Because of the remoteness of the region and the rough seas, no clean-up was mounted, and patches of asphalt-like residue stain the rocks to this day. "It looks like a cheap driveway," says David Page, a chemist at Bowdoin College in Brunswick, Maine.

In the case of the *Prestige*, the volume of oil spilt and the wildlife, fishing and tourist value of the Galician coast demanded action. The first priority in a clean-up is clear — remove oil from the beaches as quickly as possible. If washed back out to sea, or buried in the beach, oil can do more damage some other time or place.

Cleaning beaches is ideally done manually. People with shovels are the only tools sensitive enough to remove the oil while protecting the ground beneath. Only the human eye can distinguish patches of oil from the clean areas in between: on some of the Galician beaches, oil-coated rocks and apparently unaffected rocks sit side by side. And people can work on isolated rocky shores where heavy machinery cannot go.

But at Lira in early December, manual labour was having no effect. "We got the full load of a couple of tanks of the *Prestige* — 15,000 tonnes on two kilometres of coast," says Garcia. "You'd see guys working manually, and at the end of the day, the area occupied by the slick was the same."

To decide what to do next, specialists combined their experience with local knowledge in an assessment process known as net environmental benefit analysis. Factors

BOB EDMIE/AP

Volunteers clean up after the *Prestige* oil spill (far left). After the *Erika* spill off Brittany in 1999, rocks were hosed with sea water (left).

Oxford, UK, that works with the British government. The oily shore must be sheltered, otherwise the sea will wash the oil and bacteria away. Oil buried in sediments can't be digested. And if the natural bacterial growth is limited by temperature rather than by nutrients, the treatment will have no effect.

The oil must also be biodegradable: light crude oils of the type spilt by the *Exxon Valdez* are broken down easily, but heavier types contain compounds that microbes find indigestible. Finally, sites must be secluded, so that they do not offend the senses of local people and tourists during the several months the bacteria need to work. The remote Alaskan shores polluted by the *Exxon Valdez* fitted the bill³, but many others do not. "For marine spills, bioremediation is a niche market," Swannell concludes. Spanish researchers began bioremediation experiments on oiled beaches late last month, although the *Prestige's* thick, poorly biodegradable oil might not respond well to the treatment.

In Galicia, local people wanted heavy machinery to come in and remove large quantities of oil quickly. But this has its own environmental costs. There were no tracks to the worst-affected beaches and building them would have harmed the surrounding landscape. "If you've got a site that no one visits, then it's ideal to leave for natural clean-up. If it's an area that people go to all the time you can't do that, because people will be getting oil on them," says Rob Self of Oil Spill Response, a company in Southampton that worked out of the command centre in La Coruña to advise the Spanish authorities on the clean-up operation. But at Lira, oil on the surface of the beaches was likely to be washed into the pipe that supplied water to Garcia's fish farm. In the end, the desire to protect the farm tipped the balance in favour of building tracks for bulldozers and earthmovers, which scooped up the chocolate mousse from the beach.

Heavy machinery has not been used in all areas affected by the *Prestige* spill. Salt-marshes and estuaries are such delicate terrain that almost any activity does more damage than the oil. After the *Amoco Cadiz* disaster in 1978, when nearly a quarter of a million tonnes of oil were spilt off the coast of Brittany, in northwest France, heavy equipment was sent into some polluted salt-marshes, where it scraped up the top half-metre of sediment. Twelve years later, these areas had still not recovered, whereas the oiled marshes that went uncleaned seemed in good shape⁴.

In Galicia, some marshy areas were placed completely off-limits, even to people, when it became clear that volunteers were

taken into account include environmental considerations, such as whether an oiled beach is home to a breeding colony of seals or seabirds. Socioeconomic considerations also come into play: local people may rely on nearby shellfish beds, for example, or an affected beach could be a tourist destination. And practical realities, such as how much a particular clean-up option will cost, and how easy it will be to implement, are also assessed.

Every intensive clean-up option has its drawbacks. When high-pressure hot water was used to scrub the Alaskan shoreline oiled by the *Exxon Valdez* spill in 1989, beaches that got this treatment recovered more slowly than those that did not, although conditions seemed about equal after three years⁵. This technique is now used less often, says Alan Mearns, a marine ecologist with the Hazardous Materials Response Division of the US National Oceanic and Atmospheric Administration (NOAA) in Seattle, Washington. "We have a go-easy policy on using rigorous methods," he says. The coastguard understands that you don't have to go in with all guns blazing." In Spain, high-pressure water has so far been used only on man-made structures such as jetties and harbour walls.

Chemical cleaners also do damage. When the *Torrey Canyon* ran aground off the south coast of Cornwall, UK, in March 1967, oiled beaches were sprayed with 10,000 tonnes of powerful solvents and detergents, including industrial degreasers. The chemicals were more toxic than the oil: many seashore

The coastguard understands that you don't have to go in with all guns blazing.

invertebrates died, and the nutrients in the dispersants caused an explosion of seaweed growth².

Dispersants are much milder now, but they can still be used inappropriately. Peter Dyrinda, a marine biologist at the University of Wales Swansea, UK, says that after the *Sea Empress* ran aground off the southwest coast of Wales in 1996, dispersants were used on some patches of oil but not washed off. The resulting mix was more toxic than either oil or dispersant alone and killed animals that had survived until then. Such problems, as well as the fact that dispersants are ineffective against thick oil, have prevented their use in Galicia.

Softly softly

A gentler option is bioremediation, which involves using fertilizer to speed the growth of naturally occurring oil-digesting bacteria. But this won't work on every spill, says Richard Swannell, a bioremediation specialist with Momena, a consultancy company near



Common scoters heavily oiled in the *Sea Empress* disaster had a poor chance of surviving.

cleaning with excessive gusto. "People were pushing the oil into the substrate, and that has more of an effect than if you just left it," says Self. "In the end we had to close the site."

Using a combination of manual labour and heavy machinery, the *Prestige* spill has now moved from what Self calls the emergency phase — high profile, high pressure — to a long-term painstaking project. Cleaning beaches, for example, is a Sisyphean task. At Carnota, one of the most heavily affected areas of Galicia, the high-water mark on the beach in February continues to be marked by a chain of thumbnail-sized oily gobbets. Dozens of volunteers and soldiers work their way along six kilometres of sand on hands and knees, picking up the small lumps with what look like wallpaper scrapers. At other beaches nearby, people sift the sand and comb seaside plants by hand to remove the oil and stop it becoming buried.

Perhaps the most difficult decision, and one usually taken by local politicians, is when to stop. As the coastline becomes cleaner, the clean-up starts to cost more for progressively less reward and more environmental damage. On rocky shores, once the bulk of the oil is recovered, the decision often comes down to whether the beach is an eyesore. Once workers have done all they can with shovels and pompoms — balls of plastic strips that soak up oil — stubborn patches still remain under rocks and in crevices. Moving oily boulders down the beach into the surf can accelerate the natural cleaning process, as can flushing — pumping sea water over the shore and sifting the oil out of the run-off.

As the *Prestige* clean-up continues into spring and summer, researchers are starting to try to assess the spill's effects. Marine toxicologist Ricardo Beiras of the University of Vigo in Galicia hopes to produce data on damage to local fisheries on which compen-

sation claims can be based. Researchers will also be called on to pronounce when the coast has recovered. This doesn't necessarily mean that no trace of oil remains — old, weathered oil is not very toxic, and oil locked in sediments may not harm organisms.

But the definition of recovery is disputed. The owners of the *Exxon Valdez* are still fighting with Alaskans over whether the area hit by the spill is still suffering. Page, whose funding comes partly from Exxon, thinks it isn't. "With all spills there are places you can go back to and dig and find a deposit," he says. "The question is whether those isolated remnants are biologically relevant. And the answer is no, they're not."

Others disagree. Stanley Rice of NOAA's Alaska Fisheries Science Center in Auke Bay says the oil is still damaging animals that live and feed on the seashore, such as otters and salmon. "These damages are new and continuing from the remaining oil, and not just a slow recovery from the original hit," says Rice. To try to minimize such chronic effects, he advocates a swing towards more intensive cleaning. "I would push for a more aggressive clean-up, realizing that for the short term you are going to suffer more damage."

Political science

Resolving issues such as these will require more research, but studies of spills are often only carried out in the fraught atmosphere after a disaster. As a result, clean-up efforts are not as well informed or coordinated as they could be. "We don't clean up spills as well as we should be able to," says Ian White, managing director of the London-based International Tanker Owners Pollution Federation, which responds to spills around the world. "Spills take on a political significance that's hard to control."

Research tends to be improvised and opportunist, as local scientists drop what they were doing and start studying the pollution. "In the early days of a major spill things are terribly chaotic," says Dyrinda. "Scientific study is not a top priority." Like many others, Dyrinda had oil-pollution research thrust upon him when the *Sea Empress* ran aground close to where he works. Scientists in northwest Spain are having the same experience, and once again the research is not running smoothly. "There are different individual studies, but they're not coordinated. People are diverting resources from other projects," says Beiras.

But despite the gaps in our knowledge, scientists say that the major threat to the coast, should another spill occur, is not lack of research but of government preparedness. There have been six major oil spills in Galicia. The region is also a hotbed of marine science, with four institutes in Vigo alone. But such experience doesn't guarantee that government officials will talk to local experts. "There's not a lack of knowledge, there's a lack

of communication," says plankton ecologist Pablo Serret, of the University of Vigo.

Many Spanish researchers accuse their government of ignoring scientific advice in their handling of the spill — particularly in its decision to tow the *Prestige* out to sea rather than into port — and of seeking to play down the incident's severity. At the end of January, the scientific community rejected the original plan for studying the clean-up and recovery, put together behind closed doors by the government's National Research Council; the replacement was published only last week, four months after the spill. Just before Christmas, Galician researchers, worried that the government was giving Spanish researchers a bad name, put together a letter of protest⁵. They circulated it around the research community, and got 385 signatures in 24 hours.

Research into clean-up methods may be coming together. But as the strength of feeling among Spanish researchers attests, good science is of limited use unless scientists have the backing of politicians. Without that, the damage from oil spills risks going unchecked. "We haven't learnt from the past," says Beiras. ■

John Whitfield is in *Nature's* news syndication team.

1. Mearns, A. J. in *Proc. Exxon Valdez Oil Spill Symp.* (eds Rice, S. D., Spies, R. B., Wolfe, D. H. & Wright, B. A.) 309–328 (American Fisheries Society, Bethesda, Maryland, 1996).
2. *Using Oil Dispersants on the Sea* 317–318 (National Academies Press, Washington DC, 1989).
3. Bragg, J. R., Prince, R. C., Harner, E. J. & Atlas, R. M. *Nature* **368**, 413–418 (1994).
4. Baker, J. M. *Pure Appl. Chem.* **71**, 135–151 (1999).
5. Serret, P. et al. *Science* **299**, 511 (2003).

International Tanker Owners Pollution Federation

♦ www.itopf.com

NOAA Hazardous Materials Response Division

♦ response.restoration.noaa.gov



Rocks tarred with oil on the Galician coast in February.

Time for an Italian renaissance?

Italy's national science academy is celebrating its 400th anniversary, in part by cataloguing its wealth of intellectual treasures. But can it reinvent itself as a modern, politically relevant body? Alison Abbott reports.

Entering the Rome headquarters of the Accademia Nazionale dei Lincei is a quintessentially Italian experience. Housed in a Renaissance palace purchased and extended in the 1700s for the family of Pope Clement XII, the country's national science academy is located on the south bank of the River Tiber.

Visitors may climb the building's grand marble staircase without challenge from the porter. Reaching the top, you wander through a succession of empty rooms, their ceilings frescoed, their walls ornately papered and hung with paintings. The large windows, which look out over the botanical gardens of Rome's La Sapienza University, are framed by heavy brocade curtains. Only the gentle creaking of ancient floorboards breaks the silence.

Eventually, you reach the inner sanctum, where a small but dedicated secretariat runs the academy's scientific activities. "Aesthetically, it's a beautiful place to work," says one staff member. "But sometimes, when an electricity failure wipes out an hour's computer work, or a piece of ceiling falls into your coffee..." Her voice trails away, as she contemplates the frustrations of working for Europe's oldest academy, which this year marks its 400th anniversary. It's a time for celebration, but also for reflection on the Lincei's failure to exert the political influence enjoyed by its counterparts elsewhere in Europe — for which the building's crumbling grandeur provides a fitting metaphor.

The quadricentennial celebrations include the unveiling of many of the academy's intellectual treasures, which are only now being catalogued. Later in the year, for example, will come an exhibition of drawings, including a little-known sketch by Leonardo da Vinci. Already, some of the jewels in the academy's collection of historic scientific



Amid the Accademia Nazionale dei Lincei's palatial splendour, academy president Edoardo Vesentini (above) is envious of other European academies' high public and political profiles.

manuscripts have been put on display. "There are over 3,000 of them, but very few had been properly researched," says Francesca Manzari, an art historian who began cataloguing the Lincei's collection some three years ago.

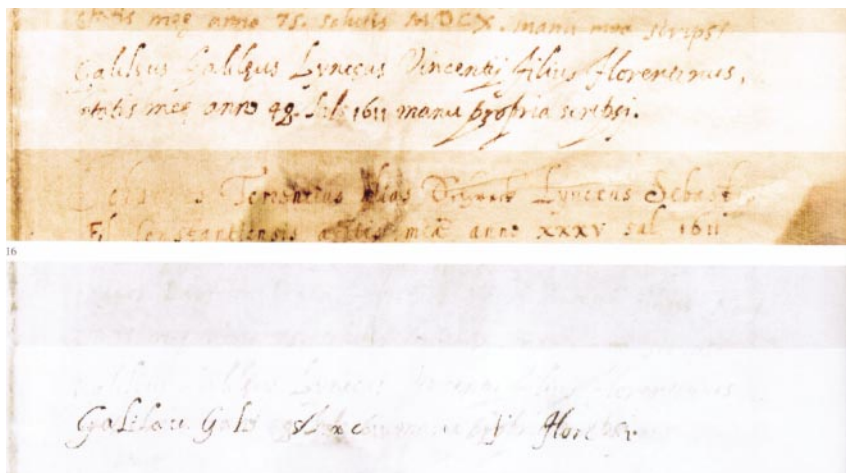
Influential ideas

The academy's members — known as Linceans — also hope that the celebrations will raise the body's profile. "No one in the streets of Italy would notice if the academy disappeared," complains the Lincei's president Edoardo Vesentini, a mathematician at the Technical University of Turin. Despite its prestige within the Italian scientific community, the academy enjoys little of the wider influence exerted by its counterparts in Britain and France, which regularly provide scientific advice to their governments.

The academy's history is a colourful one. It

was founded in 1603 as the Accademia dei Lincei — literally, the 'academy of the lynxes' — by the nobleman Federico Cesi, then just 18 years old. Cesi was an idealist who wanted to understand the world through observation — which is why he named his creation after a creature that was then thought to have particularly acute vision. That philosophy also explains the hostility encountered by the nascent academy, which immediately began to challenge received classical and biblical wisdom. Cesi was disowned by his father; other members — most famously Galileo Galilei, who became the sixth Lincean in 1611 — were denounced by the Roman Catholic Church. "The academy was born fighting," says Vesentini.

Financial and moral support for the original Lincei died with Cesi in 1630. The academy resurfaced several times, but received strong backing only in 1847 — ironically, given its



Hand of history: as part of the Lincei's quadricentennial celebrations, scientists have digitally restored the academy's original charter, featuring Galileo's signature (top, restored; bottom, before restoration).

rebellious past, from the ultra-conservative Pope Pius IX, who named the organization the Pontifical Academy of the New Lincei. With the unification of Italy in 1870, however, most Linceans swore allegiance to the new nation and its king, breaking away to form what was eventually to become the Accademia Nazionale dei Lincei of today.

By this time, many of the manuscripts and drawings of Cesi's original academy had been sold, but new collections were being assembled. And in 1883, the Italian state installed the academy in the Corsini Palace — still its home today — complete with its valuable library, on the condition that this remained open to the public.

Mergers and purges

Over the ensuing four decades, the academy enjoyed relative calm — until the advent of fascism. In 1926, Benito Mussolini established the rival Accademia d'Italia, which he housed in the Villa Farnesina, a gem of Renaissance architecture across the road from the Corsini Palace. Put under political pressure, all but eight of the 150 Linceans declared their allegiance to the fascist regime in the 1930s, and the two academies were merged in 1939. After the fall of Mussolini in 1944, however, the Lincei was re-established in its traditional form, and its membership purged of fascist appointments. It moved back into the Corsini Palace, but also inherited the Villa Farnesina, complete with treasures including a room decorated with frescoes by Rafael.

It is the academy's cultural heritage that is being emphasized in the quadricentennial celebrations. Three years ago, when Manzari and her colleagues began to catalogue the Lincei's library in preparation to exhibit its manuscripts, they faced a mammoth task. "Many manuscripts were described with only one or two handwritten sentences," says Manzari. "We found lots of surprises, lots of treasures."

These neglected riches include sacred manuscripts, plus records of major engineer-

ing and architectural works — such as accounts of payments to two of the most important baroque architects, Francesco Borromini and Giovanni Lorenzo Bernini, for projects commissioned by Pope Innocent X. "I didn't know about this resource," enthuses Jürgen Renn, a director at the Max Planck Institute for the History of Science in Berlin, who is coordinating a European project to publish documents on the Internet that reveal the evolution of mechanical knowledge.

The exhibited works also include eight volumes of coloured drawings commissioned by Cesi, who wanted to create a complete visual record of natural history. The 1,365 pages reveal the illustrators' scientific approach. Plants, for example, are shown at all stages of their development, and include details observed for the first time under the microscope — a term coined by the Linceans.

This exquisite collection — "an infinitesimal part of all the material we have", according to Vesentini — has now been digitized, and Vesentini hopes that they will be made available on the Internet. In addition, the Linceans' original charter, dating from 1611, has been digitally restored by scientists from the University of Bologna and Fotoscintifica, a digital-imaging company in Parma. On the document, the previously indistinct signature of Galileo now stands out boldly. Vesentini's dream is to annotate and digitize the library's entire contents, but its relatively meagre income in government grants means that this remains a distant goal.

"The government is not supportive," laments Vesentini, who looks enviously at Britain's Royal Society and the French Academy of Sciences — founded on Lincean principles in 1660 and 1666, respectively. Not only are these academies financed more generously, he says, but they also have a higher public profile. "Ordinary people are aware that these academies have authority," Vesentini says.

In truth, Vesentini probably overestimates the familiarity of the average British or

French citizen with their respective national science academies. But the Lincei's most obvious counterparts certainly have more political clout. "It's not clear why the Accademia dei Lincei is not a prominent player in science policy," says Royal Society president Robert May. "It's certainly not for lack of superb scientists in Italy."

The Lincei is, officially, supposed to provide scientific advice to the Italian President, but is infrequently consulted and never volunteers its opinion. In any case, the Italian presidency is a largely ceremonial office, and the academy has no meaningful relationship with Italy's real custodians of power. Successive administrations have chosen not to seek its views — even on issues such as the constant reforms that have shaken up the Italian research system over the past decade.

Some of the more active Linceans are trying to boost the academy's profile, for instance by organizing conferences for the public. "And we have started to go out into schools in different cities to give talks," says academy member Lamberto Maffei, director of the Institute of Neuroscience in Pisa, a part of the CNR, Italy's national research council.

Age-old problem

But attempts to enliven the academy have not been helped by a policy that, until recently, allowed new members to be appointed only when an existing Lincean died. Of the academy's 150 full members, 17 are in their nineties and 44 are octogenarians. Just 36 are under the age of 70. Since 2000, two or three of the oldest have been moved each year into a new category of membership, allowing replacements to be appointed from the ranks of corresponding members. This has reduced the Linceans' average age, but only from 78 to 76.

Perhaps the biggest obstacle, however, is the academy's entrenched suspicion of governments — perhaps a legacy of its turbulent history — and a curious ambivalence towards the value of assuming a higher public profile. "If people know about us but get the wrong idea about us, it could harm us," muses Vesentini. In January, for instance, one of the academy's standing committees produced a paper endorsing the continued development of genetically modified crops. Tellingly, however, this was published as a statement by the committee, rather than by the academy as a whole.

Some of the academy's new, younger members argue that it's time for these attitudes to change — although they're not optimistic. "The Accademia doesn't push hard enough to get things done," says Jacopo Meldolesi, a molecular neuroscientist at the San Raffaele Scientific Institute in Milan, who was elected to the Lincei's ranks last year. "The membership is too old."

Alison Abbott is *Nature's* senior European correspondent.

♦ www.lincei.it/english.version/default.english.html

Max Planck: closures will damage German science

Cosmochemistry and geochemistry departments should continue their innovative work.

Sir — Your News story “Max Planck plans double blow to chemistry” (*Nature* **422**, 105; 2003), reporting that the Max Planck Society is seriously considering closing its departments of cosmochemistry and geochemistry in Mainz, is most disturbing. Insofar as the Max Planck Society is concerned with pursuing cutting-edge scientific research programmes, such action is difficult to understand.

Research in cosmochemistry is a vital laboratory activity, linked with astrophysics and observational astronomy. It has active and strong connections to study of the early chemical evolution of the Universe, cosmology, and the Solar System’s formation and evolution. Important pioneering work, both in terms of new discoveries and the development of new generations of instruments, is being

carried out under the direction of Günter Lugmair and is the focus of widespread and active interest. Cosmochemistry is an intellectually and technically exciting area and will continue to be so during the next few decades.

The area of terrestrial geochemistry has grown enormously in terms of the depth of understanding of the chemical processes related to Earth’s structure and dynamics and of the geophysics of Earth’s thermochemical and dynamical engine. There are important new possibilities in studies of environmental geochemistry, both in measurements and in theory, that are key to understanding the migration of elements, not only in Earth’s deep interior, but also in its near-surface aqueous environments. A whole new array of techniques and instrumental approaches

are now being developed, some being pioneered at Mainz under the leadership of Albrecht Hofmann.

The absence of these research areas from the Max Planck Society’s science programme will, in our view, have a devastating effect on the level of science in Germany and will be negative for the rest of Europe. There is no doubt that some restructuring of German research is needed for fiscal and management reasons. However, closing leading institutes that are doing vital and innovative research is an action that we must deplore.

G. J. Wasserburg*, D. P. McKenzie†

*USA Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena California 91125, USA

†Department of Earth Sciences, University of Cambridge, Cambridge CB3 0EZ, UK

Max Planck: cuts were decided years ago

Sir — In your News story “Max Planck plans double blow to chemistry” (*Nature* **422**, 105; 2003), you state that two departments at the Max Planck Institute for Chemistry in Mainz are the “first casualties of a budget freeze” at the Max Planck Society. This is not correct.

The closure of one department was decided years ago and has nothing to do with the present budget constraints of the Max Planck Society. The federal consolidation programme, aimed at reducing staff numbers in former West German institutes, has obliged the Max Planck Institute for Chemistry to cut 25 staff positions. This includes eliminating Professor Günter Lugmair’s position when he retires in 2005.

The institute’s further development has to be seen within the context of the overall development of the Max Planck Society. Within this context, the institute is free to set its own scientific priorities, which include the continued use of the ion microprobe pictured in your story.

Bernd Wirsing

Max Planck Society Press Office, Postfach 10 10 62, 80084 München, Germany

Our reporter contacted the Max Planck Society’s press office as well as the heads of departments at the Institute of Chemistry in Mainz. None of these told him that the closure of the cosmochemistry department had been decided several years ago — Editor, *Nature*

Leading edge lasers

Sir — Your News in Brief story “Spending spree gives German physics a high-energy boost” (*Nature* **421**, 682; 2003) reported that free-electron lasers generate intense X-ray radiation for the study of molecules during chemical reactions. In fact, these lasers have operated from the microwave to the vacuum ultraviolet and can be used to study not only molecules but atoms, condensed matter, plasmas and even elementary particles via Compton back-scattering. Thanks to their performance advantages, which include tunability, pulse brevity, high repetition rate, capacity for pump–probe synchronization with other light sources, and high average and peak powers, the development of free-electron lasers is being advanced on many fronts.

Swapna Chattopadhyay

Thomas Jefferson National Accelerator Facility, 12000 Jefferson Avenue, Newport News, Virginia 23606, USA

Vigilance is vital to avoid conflicts of interest

Sir — Rahul K. Dhanda makes an excellent point in his Correspondence “Time for bioethics and business to start talking” (*Nature* **421**, 573; 2003): the separation between bioethicists and biotechnologists constitutes a risk to both groups, as well as to the public good. He overstates, however, the potential for synergy.

The rift between bioethicists and the

biotech industry is not just cultural, as Dhanda suggests, but also functional. Bioethicists have a social role that is sometimes at odds with the goals (and social role) of industry. The public expects bioethicists to offer independent, reasoned commentary on complex issues in health-care and the life sciences. Nowhere is such commentary more crucial than in the rapidly expanding realm of biotechnology. Surely Dhanda would agree that unreflective collaboration with industry would hinder the ability of bioethics to meet this expectation.

It is true that bioethicists have often been preoccupied by questions of how they can maintain their integrity if working with industry. But what Dhanda calls “endless discussions over conflicts of interest” are, I fear, a necessity. Bioethicists need to carry on this discussion, but as well as including industry they must also try to determine what sorts of engagement with industry are least likely to corrode public trust in bioethics, and what kinds of structural safeguards are possible.

The Enron scandal taught us that the accounting profession must become more vigilant in its struggle to avoid conflicts of interest, and indeed may need to consider significant structural reforms if the profession is to retain its useful role in the functioning of financial systems. Similarly, bioethicists should not stop talking about conflict of interest: vigilance on this count is the discipline’s only hope of continuing to merit the public’s trust.

Chris MacDonald

Department of Philosophy, Saint Mary’s University, Halifax, Nova Scotia B3H 3C3, Canada

Bringing down the barriers

Public communication should be part of common scientific practice.

Jaap Willems

The public is fascinated by science, particularly astronomy. But despite most researchers recognizing the necessity of communicating to the public, many of them fail to do so. Although the media is the main source of scientific information for most people, scientists throw up barriers to their work being publicized. Scientists need to popularize their subject as, sooner or later, society will have to deal with the results. Not only do people need to keep up to date with rapidly changing knowledge, but ignorance often leads to fear.

Although some scientists accept that the public must be kept informed and interested if they are to obtain funding, many are puzzled by the suggestion that the popularization of, for example, chemistry is important for creating public support. Surely science no longer needs to justify itself, they ask? Furthermore, many researchers would — quite wrongly — treat with derision the idea that scientists need to popularize their work if they are to reach fellow professionals in their own or related fields. Yet various surveys have revealed that communication between fellow professionals often takes place through the mass media.

Most public communication about science is channelled through daily newspapers, special-interest magazines and television. In the Netherlands, articles written by researchers themselves are occasionally published in newspapers or in popular science magazines such as *Natuur & Techniek* and *Greenpeace*. However, about 90% of these articles are written by science journalists, most of whom do not have scientific qualifications. And according to surveys in the Netherlands, Germany and the United Kingdom, the public is dissatisfied with the media's reporting of innovations in science and technology. Media reports can heighten public fear of certain areas, for example biotechnology, according to Eurobarometer (http://europa.eu.int/comm/public_opinion).

A different approach is needed. Science communication professionals have long advocated a shift from the one-way channel of the mass media, towards interactivity — science and discovery centres, public lectures and company or institution open days — to bring researchers into direct contact with the general public. If nothing else, the resultant dialogue is a useful addition to media reporting in conveying accurate information and reducing fear of new technologies.

But scientists must also find ways of improving communication through the



Revolution: to capture public support, scientists must smash the obstacles between them and the media.

media, as this is familiar to many and is the most efficient way to reach large numbers of people. Yet our survey (see footnote) reveals barriers to such communication, such as unreasonable demands from researchers that journalists' reporting must be full and complete, or the lack of appropriate expertise by journalists — ignorance of basic technical terms, or a desire to sensationalize or exaggerate the discovery.

Management and public-relations (PR) departments frequently block contacts between scientists and the media. Our survey indicates that only one-third of researchers in the Netherlands can decide what they tell journalists. The rest have to defer to managers and PR departments, even in universities. The PR department initiates contact with the press. Of course, PR officials have a better understanding of the media and more contacts than scientists. Nevertheless, many Dutch scientists do not want to help PR departments popularize their research as they would prefer to do it themselves.

PR officials, of course, are usually only interested in good news about the research in their institutions. Journalists are more interested in bad news (such as risks associated with genetic modification) and would prefer to publicize details before the full work is published in scientific literature. These separate, selective agendas provide further barriers to the communication of science.

That 90% of scientists in our survey believe that a journalist's reporting should be full and complete, and the journalists should allow the scientists to check their story and make requested changes before publication betrays an ignorance of journalistic methods. As journalists would naturally not agree to these conditions, scientists are very reticent about cooperating with the press. Virtually none of our respondents knew the names of

the science editors of the major Dutch quality newspapers, many of whom have been writing about science for years.

Finally, almost half of our respondents had never written an article for a wider general readership, while a further 40% did so only very rarely. Only 10% regularly write articles about their own speciality for a general readership, a fraction that included a disproportionate number of ecologists writing about environmental issues. Although not every scientist can be expected to write popular and/or accessible articles about their work regularly, and the media could not handle the resultant volume of material, the fact that so few biologists take an active part in popularizing their work highlights, once again, their lack of interest in public communication.

Many scientists, used to writing scientific articles, lack the rather different writing skills needed to bring their work to a wider audience. In addition to this, many feel that popularization would reduce their status among their peers. Yet almost every university offers courses in science communication, and although scientists go on these courses, they are generally regarded as being on the margins of university education. If we truly want the media to expand and improve its coverage of science and technology, more researchers need training in public communication and must be prepared to use these skills by participating in public events, writing popular, accessible articles, and cooperating constructively with science journalists. ■

Jaap Willems is in the Department of Science Communication, Vrije Universiteit Amsterdam, 1081 HV, The Netherlands. Further information available in *Biologen en Journalisten* (Biologists and Journalists) by Jaap Willems, Betteke van Ruler, Linda Hartman and Neil van der Veer (Enschede, Amsterdam, 2002). See also www.bio.vu.nl/WillemsinNature.pdf.

A word in your ear, ambassador

How can scientists provide more effective advice to the United Nations?

Knowledge and Diplomacy: Science Advice in the United Nations System

by the Committee for Survey and Analysis of Science Advice on Sustainable Development to International Organizations, Development, Security, and Cooperation, National Research Council
National Academies Press: 2002. 120 pp.

\$28 (US), \$33.75. Available free online at: <http://www.nap.edu/books/0309084903/html>

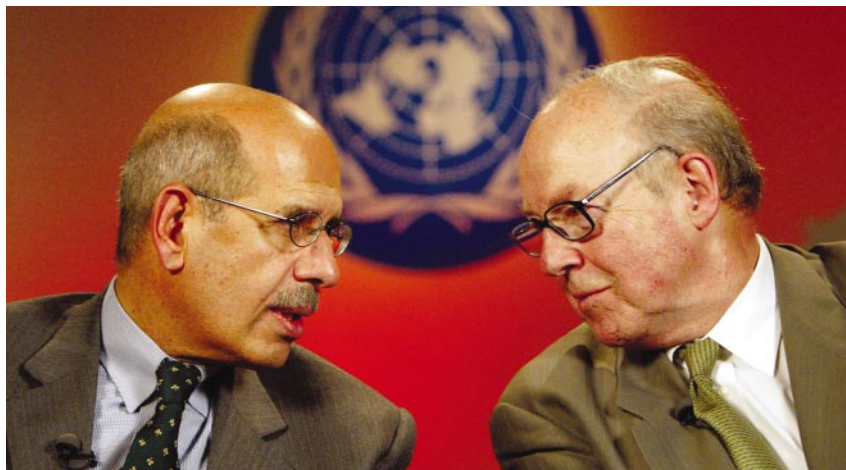
Ehsan Masood

It isn't often that the arcane world of science advice in the United Nations (UN) system makes front-page news all over the world. But science at the UN has lately found a mass audience thanks to Hans Blix, its chief inspector for biological and chemical weapons in Iraq, and his frequent reports to the UN Security Council.

Blix spent many years as director general of the International Atomic Energy Agency (IAEA) in Vienna, Austria, a UN agency that was created by President Eisenhower to monitor and oversee the peaceful use of nuclear power. The IAEA is perhaps the only significant omission from *Knowledge and Diplomacy*, which is an otherwise excellent survey of the short history and practice of science advice within the UN system. The report was produced at the request of the science adviser to the US secretary of state Colin Powell. He asked the US National Academy of Sciences to survey and analyse science advice to the UN agencies that are involved in international sustainable development, paying close attention to water, fisheries, oceans and energy.

Some observers may well ponder the merits of such a request from the present US administration. This government has after all chosen to reject the recommendation of Mohamed El Baradei, the present head of the IAEA, that the UN weapons inspectors be allowed more time in Iraq. The administration has similarly had difficulty in accepting international scientific opinion on the merits of the Kyoto Protocol, and has yet to sign up to the UN Convention on Biological Diversity.

The IAEA reports to the Security Council, which, strictly speaking, is not directly involved in international sustainable development. But an analysis of the agency, which is effectively the only permanent source of science advice to the UN's most powerful organ, would have made this report more complete. Furthermore, the IAEA itself is a vocal advocate of nuclear energy as the world begins to look for more sustainable alternatives to fossil fuels.



IAEA chief Mohamed El Baradei (left) confers with UN chief weapons inspector Hans Blix.

Knowledge and Diplomacy makes several recommendations that are aimed principally at giving science and scientists greater influence in the UN system. The recommendations include calling on the UN to commission more independent, peer-reviewed scientific advice; make greater use of comprehensive scientific-assessment methods and institutions, such as the Intergovernmental Panel on Climate Change (IPCC); strengthen scientific advisory capabilities within member states; and adopt a transparent set of procedures for all UN science advisory work. But top of the list is a recommendation to establish a science adviser's office attached to each UN organization that has "substantial responsibilities" for implementing sustainable development, including the office of Secretary General Kofi Annan.

The report envisages that a science adviser would help each organization, and the UN as a whole, to recognize policy issues that would benefit from science advice; assist in drawing up a list of questions to ask; commission relevant work from external organizations; interpret research findings; and outline possible policy implications from the advice being given.

This recommendation, although admirable in its scope, is likely to be opposed by many UN member states. There are several reasons for this. Perhaps foremost is the assumption that the science advisers and their staff will be immune from political influences, and will be allowed to ask questions and chart an independent programme of research free from political interference. With the best will in the world, this rarely happens in the UN system. Every paid official in a UN centre is subject to the most intense lobbying from governments, non-governmental organiza-

tions and industry. A new science adviser's post will be no different, so the advice given can never be truly independent.

UN member states will fear that a science adviser's office attached to Kofi Annan will be answerable to the Security Council, which could influence both its programme of work and the sources of information used in its reports. Many UN member states will oppose such centralization of science advice, not because they particularly favour independent advice but because they will have less influence on the kinds of questions that such an adviser will be allowed to ask, and the experts that he or she will be encouraged to consult.

A further complicating factor is that many of the UN's sustainable-development organizations have scientific advisory mechanisms already in place, and it is unclear how these will work with a new layer of more centralized science advice. Some of these mechanisms are deeply politicized, but a few are of the type advocated elsewhere in *Knowledge and Diplomacy*.

For example, the UN Environment Programme (UNEP) has formally contracted its scientific advice to an external body, the World Conservation Monitoring Centre in Cambridge, UK. Others, such as the UN Convention on Biological Diversity, receive mostly in-house advice from a committee of experts appointed by the governments who ratified the convention. The UN Framework Convention on Climate Change by contrast, takes its advice from the vast and comprehensive IPCC.

Far from giving science a stronger voice in the UN system, an additional layer of unsolicited science advice could have the opposite effect. History tells us that unless advice has been asked for, it is often ignored or rejected.

D. GUTTENFELDER/AP

In 1995, UNEP published an impressively weighty document called the *Global Biodiversity Assessment*. It was designed to provide its sister organization, the Convention on Biological Diversity, with the latest information on the state of the world's threatened species. But this document was ignored by the governments who belong to the biodiversity convention. This was partly because the assessment had not been asked for, but also because it more or less ignored the plight of the often very poor people who live in or near the world's biodiversity hotspots.

Many of the UN organizations involved in the *Global Biodiversity Assessment* have now set up what is called The Millennium Ecosystem Assessment. This ambitious undertaking, involving governments, non-governmental organizations and industry, seeks to review the state of the world's ecosystems and promote people-centred conservation policies. Interestingly, this is an example of another of the recommendations from *Knowledge and Diplomacy*: making better use of scientific assessment facilities, such as the IPCC. Perhaps this is the recommendation to which the US secretary of state ought to pay most attention, particularly if the aim of *Knowledge and Diplomacy* is to build greater trust in science and scientists. ■

Ehsan Masood is at Leadership for Environment and Development International, based at Imperial College, 48 Prince's Gardens, London SW7 2PE, UK, and is a former science writer at Nature.

The barnacle years

Darwin and the Barnacle

by Rebecca Stott

Faber and Faber: 2003. 336 pp. £14.99

Frederick R. Schram

Books about Charles Darwin now form a considerable library in their own right.

Yet one period of his life seems to have been rather neglected: the years between 1844, when he completed the first draft of his ideas about the



Robert Grant: an early influence on Darwin.

species problem, and the publication in 1859 of *On the Origin of Species*. These are the barnacle years, when Darwin trained himself in zoology.

In *Darwin and the Barnacle*, Rebecca Stott tries to throw light on what happened during this period. She tells us that "it is the story of one tiny creature and history's most spectacular scientific breakthrough". I suppose it is difficult to avoid hyperbole when it comes to Darwin — after all, he wrote some 19 books, many of them founding canons of their individual sciences. Few scholars have been so productive and so influential in so many fields. The tendency to canonize Darwin is almost irresistible, especially for someone who is English and a denizen of Cambridge, as is Stott.

This book focuses on the eight years between 1846, when Darwin published his last geology tome, and 1854, when the final barnacle volume appeared. According to Stott, Darwin complained in his letters that this was a long time. But I doubt that any taxonomist starting work on a group from scratch would find any need to apologize for taking eight years to produce four monographs comprising over 1,200 pages. In fact, throughout his career Darwin learnt quickly and was a fast worker. He put his species theory aside because there were things he needed to clarify in his own mind, and he recognized the need for hands-on experience with real species.

'Mr Arthrobalanus', a specimen of boring barnacle picked up by Darwin on the coast of Chile in 1835, keeps appearing again and again throughout the book, like the tolling of a funeral bell. If Stott intended this as a dramatic device to convey the unending tedium

of Darwin's barnacles, she has succeeded. By mid-volume one cries "Enough already!" But she adds insult to injury by never revealing what Mr Arthrobalanus was (apart from its eventual scientific name, *Cryptophialus minutus*). A hint is given in the footnotes to chapter 11, but if you don't read footnotes at the back or are not familiar with barnacles then you are out of luck. Mr Arthrobalanus is not a true barnacle, but rather a member of an entirely different superorder.

There are also errors and omissions. Mid-way through, Stott takes up the cause of the peculiar stalked barnacle *Ibla*. Unlike most hermaphroditic barnacles, *Ibla* has a unique way of separating the sexes, with tiny males that are little more than gonad bags attached outside the female's shell. But Figure 14, which purports to illustrate the vexing *Ibla*, actually shows a scalpellomorph, which represents a completely different suborder. We are also introduced to *Proteolepas*, but we are not told that this mysterious animal proved to be nothing but a parasitic isopod.

Nevertheless, there is much to enjoy in *Darwin and the Barnacle*. The description of a Victorian water cure is delightful. Darwin was addicted to them, and the episode at Dr Gully's establishment at Malvern is told with the right degree of tongue in cheek. The story of the last months of Darwin's daughter, Annie, is incredibly touching, linked as it was with a hopeless water cure for the dying girl. And the chapter on Robert Grant, the Darth Vader of sponge studies, working by night in his "castle by the sea" had me wishing for more. Stott tells us too of young Thomas Huxley on the make — not a pleasant person at all, by the sound of it — and poor John Coldstream, wrestling with his sexuality.

What bothered me most about the book, however, is that Stott seems to have missed the point about the barnacle years. This was when Darwin really came to grips with variation in species. He knew he had to do this. He went looking for variation in nature — his species theory demanded it — and by golly he found it. The problem is that he overestimated it. What Darwin recognized as species are now, more often than not, classified as genera, and what he delineated as varieties are today separate species.

For an accurate scientific account of Darwin and barnacles, I recommend William Newman (*Crustacean Issues* 8, 349–434; 1993). But if you want to meet some interesting characters, *Darwin and the Barnacle* is great. ■

Frederick R. Schram is at the Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Post Box 94766, NL-1090 GT, Amsterdam, the Netherlands.

Television review

Walking tall, not talking heads

Walking with Cavemen

Executive producer and director

Richard Dale

A BBC/Discovery Channel/ProSieben co-production

Clive Gamble

How many TV series can claim that all human life is here? *Walking with Cavemen*, the third outing for the BBC's immensely popular format of wildlife documentary, computer-generated images, animatronics and stunning locations, can do just that. It ranges in time from 4 million to a mere 30,000 years ago. In four 30-minute programmes, the first of which was aired in Britain on 27 March on BBC1 and which will be shown on the Discovery Channel in June, it shows us nine hominid species, including ourselves, doing what we have always done best: eating, flirting, mothering, shouting and banging rocks and heads together.

Actors play the cast of hominids. At times it seems we have strayed into the UK House of Commons. There is the 'quiet one', *Australopithecus afarensis*, and a very excitable, even reckless, *Homo habilis* with a splendid set of Victorian sideburns. They scream at each other from different trees in

the savannah while the docile vegetarians *Paranthropus boisei* lie deep in the reeds, where they become extinct.

Presenter Robert Winston sums up the aim of the series: "By watching their everyday lives closely we can see the seeds of humanity developing in these apemen and also we can see how much of the apeman remains alive in us." *Walking with Dinosaurs* was criticized for mixing up fact with speculation, particularly the sounds and colours of the beasts, for which there is scant evidence. The inclusion of Professor Winston as a Dr Who of human evolution, travelling through time to meet our ancestors, neatly defuses this charge. Instead we are allowed to reflect on our fragile humanity by visiting our distant relatives as they played out their equally small lives without the encumbrances of modern living.

Winston closely observes these unruly ancestors. He even tells a joke to a Neanderthal family. Their silence is stony. The point of the joke is to show that their brains were not yet wired for imagination. A brave scientific assumption as they had spent the morning pushing boulders over cliffs onto unsuspecting mammoths.

Of course I can quibble about details. Why was *Homo ergaster* white- not black-skinned? Couldn't Neanderthals make themselves a decent pair of boots? Monogamy has never been the norm, and language could be more recent than is depicted. But such criticisms miss the point. Scheduled for BBC1, a mainstream channel, this series is not about the debate over human evolution where the 'probablys' and 'how do you know?' are expected. That is for more specialist channels, such as BBC2 and Channel 4. This series is about drawing in a new and much larger audience to the scientific excitement and contemporary importance of asking where we came from. The programme's makers have succeeded precisely because the series' driving force is drama rather than debate.

How we represent the past, especially the deep past, has always been problematic. Stephanie Moser, in her excellent dissection of the iconography of human origins (*Ancestral Images*, Sutton Publishing, 1998), reminds us that the familiar images of cave-men are ways of expressing arguments that are difficult to put into words. Archaeologists have always borrowed from much older traditions of depicting outsiders, such as club-wielding wildmen, to make our ancestors scientifically believable. *Walking with Cavemen* provides a new slant on those vital, visual arguments. Raquel Welch and her million-year-old dinosaurs, Stanley Kubrick and the apemen of 2001, eat your heart out!

Clive Gamble is at the Centre for the Archaeology of Human Origins, University of Southampton, SO17 1BJ, UK. His book *Timewalkers: The Prehistory of Global Colonisation* has recently been reissued by Sutton Publishing.

Practical proteomics

Proteins and Proteomics:

A Laboratory Manual

by Richard J. Simpson

Cold Spring Harbor Laboratory Press: 2003.

926 pp. \$250 (hbk), \$185 (pbk)

Stephen Oliver

We now have the complete sequences of a large number of genomes at our disposal. Each of these provides us with a complete inventory of the working parts of an organism's cells. The difficult bit is discovering what all these working parts do. Functional genomics is all about the analysis of gene action and interaction on a genome-wide scale. Four levels of analysis are commonly exploited: genes (the genome), messenger RNA (the transcriptome), proteins (the proteome) and metabolites (the metabolome). Although the genome is relatively constant, the complement of the other 'omes' changes with the physiological, developmental or pathological state of the organism.

Proteomics is probably the most important, but it is also the most difficult to study in a comprehensive manner. One reason for this is that the proteome is multidimensional: a single gene can specify a number of different protein products that may themselves be modified post-translationally by the covalent attachment of a range of functional groups. Moreover, proteins commonly occur in complexes whose composition and conformation have a profound effect on both their activity and their stability. Proteomics attempts to grapple with this complexity by using a vast range of molecular, chemical and genetic techniques.

This superb volume embraces this diversity and, even if proteomic techniques are far from comprehensive in their scope, this weighty tome is certainly comprehensive in its description of them. Everything is here, from detailed laboratory protocols to clear explanations of the theoretical basis of the methods used. The book is as up-to-date as possible. It is peppered with references from 2002, and deals fully with recent technical advances, such as protein microarrays.

Richard Simpson even manages to slip in bits of history, such as the derivation of the name Coomassie for the blue dye used to stain proteins in gels. (It was named to commemorate the British conquest of Kumasie — the capital of the Ashanti, in what is now Ghana — and was originally used to dye woollen jumpers.) Such nuggets make this volume more than just an excellent laboratory manual; it is also a book to dip into and enjoy.

Stephen Oliver is in the School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK.

BBC



Metamorphosis: *Homo sapiens* gets a makeover as *Homo ergaster*.

Dynamic diversity

Sandra Knapp

In the aftermath of the Earth Summits in Rio de Janeiro in 1992 and in Johannesburg in 2002, conservation has become inextricably linked with sustainability. The United Nations Convention on Biological Diversity (CBD) enshrines this linkage in its objectives: the conservation of biological diversity; the sustainable use of its components; and the fair and equitable sharing of its benefits. The CBD defines sustainable use as the “use of the components of biological diversity in a way and at a rate that does not lead to the long-term decline of biological diversity, thereby maintaining its potential to meet the needs and aspirations of present and future generations”. Interestingly, however, conservation itself — the central plank of the convention — is left undefined. Perhaps this is because we all feel that we know what conservation means, and that its definition is superfluous. But do we really know what we are talking about?

According to the *Oxford English Dictionary*, conservation is “the action of conserving; preservation from destructive influences, natural decay or waste; preservation in being, life, health, perfection, etc. or preservation of existing conditions, institutions, rights, peace, order, etc.”. All quite

straightforward until one looks a little bit closer at just what this concept means for us as just one of many millions of species, most of which are still unknown, that share the planet Earth. *Homo sapiens* left Africa less than 100,000 years ago, colonizing all habitable continents and causing the extinction, directly or indirectly, of other hominid species such as the Neanderthals. Our species was (and still is!) an invasive mammalian weed, monopolizing and changing forever the habitats in which it becomes established.

So just what do we want to conserve? Conservation seems to imply stasis — in conserving a species, habitat or lifestyle, we expect it to remain the same. But we all know that conditions change — our world today bears little resemblance to that inhabited by *Tyrannosaurus rex* or the woolly mammoth. So what should we conserve? The situation as it was yesterday, last week, a decade ago? A world in which the rich stay rich and the poor stay poor? Even the conservation of biological diversity is fraught with questions — do we want to preserve species, such as the panda, that appear to be headed for extinction unaided by humans, or is it more important to conserve habitats, in which species are left to flourish or perish naturally?

None of these questions has an easy answer; this is partly why conservation, and particularly the conservation of biological diversity, is such an emotive issue. But it is emotive for another reason too: humans, like members of other species, are astoundingly egocentric. We see everything in terms of ourselves — what can biodiversity do for me? What benefits can we obtain from its use and, by extension, its conservation? But humans have gone even further; our alienation from nature is so complete that many of us do not think of ourselves as dependent upon the world around us, nor even as biological entities, despite abundant evidence of our increasingly devastating impact on the planet: deforestation, wildfires, falling water tables and catastrophic air pollution.

This apparent duality — wanting to conserve biological diversity, while continuing to increase our consumption of the Earth's resources — cannot continue. William Rees has said that the average human ‘ecological footprint’ — the area of productive land and water needed to support that human — measures 2.3 hectares, whereas the world contains only 1.9 hectares per person. Yet the human population is still growing. If these calculations are correct, our use of our planet is clearly not sustainable and we certainly do not want to conserve the situation in which we find ourselves today.

Conservation

Preserving nature is not about stasis, but about maintaining the exciting, ever-evolving variety of life on Earth.

In defining sustainable use and linking it to conservation, the CBD has set an almost impossible task — how do we know if we are consuming resources too rapidly? How can we tell if the potential for the as-yet-unknown aspirations of future generations will be met? The economist Herman Daly has suggested that sustainability is better defined as throughput from and back to nature, and thus a non-declining throughput can provide a measurable standard for conservation. But how should throughput be measured?

Nature is not a machine — always punching out metal parts of the same size and shape. The natural world has some great advantages over machines, such as resilience and interconnectivity. The natural world is a dynamic, ever-changing place, even in the absence of human intervention. The challenge is to preserve the dynamism of a natural world in which humans are relative newcomers.

Effective conservation of both resilience and interconnectivity will require the input of science. Without an understanding of which species live in which habitats, how they interact, their relationships through both ecological and evolutionary time and how they are affected by human activities, how can we possibly do anything other than fence off great portions of the globe and hope for the best? Conservation of a dynamic, resilient natural world will require a great deal of cooperation and an immense synthesis of existing information.

The choice is stark: we can work hard to conserve a dynamic natural world of which we are an integral part, or we can fail and be faced with the equivalent of a white room wallpapered with photographs of the species and habitats with which we used to share our planet.

Sandra Knapp is in the Department of Botany, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

FURTHER READING

Raven, P. H. *Science* **297**, 954–958 (2002).
Rees, W. E. *Nature* **421**, 898 (2002).
Daly, H. E. *World Bank Annu. Rev.* 2002 15–17 (2002).
Orr, D. W. *Conserv. Biol.* **16**, 1457–1460 (2002).
O'Riordan, T. & Stoll-Kleeman, S. (eds) *Biodiversity, Sustainability and Human Communities* (Cambridge Univ. Press, 2002).
Royal Society Working Group Report: Measuring Biodiversity (R. Soc., London, 2003).

HISTORICAL PICTURE ARCHIVE/CORBIS



Life's rich tapestry: but nature doesn't paint a static picture, so we shouldn't try to preserve one.

Mixing with latitude

Chris Garrett

Measurements over a range of latitudes support a theory relating ocean turbulence to internal waves. The upshot should be easier mapping of ocean mixing, and eventually better climate models.

Anyone who looks at the sky can see that clouds have a big effect on weather and climate. But global computer models cannot deal explicitly with individual clouds, let alone the water droplets in them. Phenomena that mix the ocean are likewise important but too small to be resolved in models of ocean circulation. These 'subgrid-scale' processes need to be mapped, in both the atmosphere and the ocean. They also need to be understood so that their effects can be 'parametrized' — that is, represented in terms of larger-scale features that are dealt with explicitly in the models. Only in this way can we correctly allow for changes in such processes and their effects in a changing world. As they describe on page 513, Gregg and colleagues¹ have produced an exciting advance in ideas about how both mapping and parametrization in the ocean might be accomplished.

The overall oceanographic context for their observations is as follows. Below a uniform surface layer that is typically some tens of metres thick, and directly stirred by wind and surface cooling, the ocean is stratified according to density, with lighter water above denser water. This gives it stability, which from time to time is disrupted by bursts of turbulence and mixing. The mixing helps to redistribute ocean properties, influencing circulation and heat transport, and hence climate². Mixing also redistributes nutrients and dissolved gases, thus affecting biological productivity in the oceans.

Most of the turbulence and mixing is caused by an 'overturning instability', which has a vertical scale that varies from a fraction of a metre to many metres and is associated with the vertical gradient, or shear, of transient horizontal currents. These currents are a manifestation of internal waves (Fig. 1) — fluctuating motions that resemble waves at the ocean surface in relying on restoring buoyancy forces associated with a vertical change in density, but having much longer periods (typically tens of minutes to many hours). Some of the internal-wave energy originates as internal tides³, which are waves of tidal period generated by the main tidal currents flowing over bumps on the sea floor. Another source is the 'inertial' waves that are generated as the ocean responds to currents driven at the sea surface by the winds of fast-moving storms⁴. These waves have a period, related to the Coriolis force resulting from

Earth's rotation, of 12 hours at the poles and much longer at low latitudes.

As the internal waves propagate into the ocean interior, some energy spreads into waves with other periods and generally smaller vertical scales, causing a greater vertical shear and hence more likelihood of wave-breaking, turbulence and mixing (Fig. 1). This cascade to smaller scales is reminiscent of what occurs in turbulent motions in unstratified water, with big eddies tearing one another apart and giving rise to ever smaller eddies, and it is also described by nonlinearities in the governing equations. The processes are, however, subtle and not fully understood. Nonetheless, a plausible model^{5,6} suggests that the cascade is not only more vigorous in places where internal waves are more energetic, but is also a function of latitude, with a slower cascade as one approaches the Equator.

It is this prediction that is so nicely confirmed by Gregg *et al.*¹. They have simultaneously measured the energy level of internal waves (as largely represented by the shear of currents on a vertical scale of the order of ten metres) and the turbulent mixing at scales of a centimetre or less. They reinforce their earlier confirmation⁷, from a limited range of latitudes, that turbulence depends on the internal-wave energy, but they now have enough data from low latitudes to show that, as predicted, a given energy level at low latitudes causes much less mixing.

This confirmation of theory is as if a cloud physicist had shown that it is no longer necessary to study individual droplets, but only the larger-scale features of clouds. In the ocean we may no longer need to undertake the technically demanding measurement of the turbulence itself, but can predict it from motions at much larger scales. To be sure, the theoretical underpinnings of the formula are not completely robust, and it is known to break down in some locations⁸. We need more theoretical studies⁶ and numerical simulations⁹ of the energy cascade of internal waves. But it does seem that we have a good provisional basis for indirectly mapping the global distribution of ocean mixing, and of any seasonal or longer-term variations.

Such mapping would still require vertical profiles of horizontal current. This is a more sophisticated measurement than is routinely made. Alternatively, improvements in measuring and analysing profiles of temperature

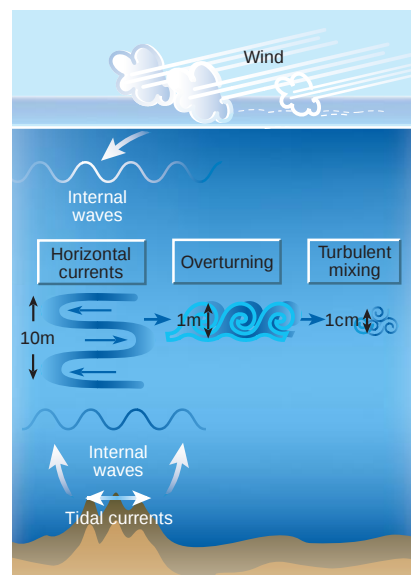


Figure 1 Internal waves and scales of ocean mixing. Waves produced at the sea surface and sea floor propagate into the ocean interior, generating small-scale mixing processes that affect circulation, heat transport and nutrient distribution (and so biological productivity). In confirming theory, two conclusions emerge from the observations of Gregg and colleagues¹. First, that the vertical gradient of the horizontal currents with vertical scales of tens of metres is a good guide to overturning at a scale of a metre or less and turbulent mixing at the centimetre scale. Second, that the connection is strongly dependent on latitude. Note that the scale of this graphic is distorted to show processes at scales of metres and less in an ocean that is about 4,000 m deep.

and salinity alone, from ships or profiling floats¹⁰, could provide a simpler approach based on analysis of, say, the overturning instability itself¹¹.

Even with a confirmed parametrization of mixing, predictive models of ocean circulation will still need information on the internal-wave field, probably requiring a model to run in parallel. In a windier world, for example, internal waves might be more energetic but lose more of their energy closer to their source, changing the strength and spatial patterns of mixing rates. This kind of feedback on circulation and climate will have to be allowed for. Finally, although mixing from internal-wave-breaking is a key

phenomenon in the deep ocean, particularly over rough topography, it may actually be less important for changing water properties in the top kilometre or so — there, some combination of air–sea interaction and stirring by eddies, with scales of tens or hundreds of kilometres¹², may dominate.

There is still much to do before we can claim to have simple operational techniques for monitoring ocean mixing rates, and a full understanding of the mechanisms and effects of mixing. But the connections between motions with scales from millimetres to thousands of kilometres are becoming clearer, and we are closer to having the parametrization of small-scale processes that is needed if models are to have predictive capability.

Chris Garrett is in the Department of Physics and

Astronomy, University of Victoria, Victoria, British Columbia V8W 3P6, Canada.

e-mail: garrett@phys.uvic.ca

1. Gregg, M. C., Sanford, T. B. & Winkel, D. P. *Nature* 422, 513–515 (2003).
2. Rahmstorf, S. *Nature* 421, 699 (2003).
3. Egbert, G. D. & Ray, R. D. *J. Geophys. Res.* 106, 22475–22502 (2001).
4. Alford, M. H. *J. Phys. Oceanogr.* 31, 2359–2368 (2001).
5. Henyey, F. S., Wright, J. & Flatte, S. M. *J. Geophys. Res.* 91, 8487–8495 (1986).
6. Müller, P., Holloway, G., Henyey, F. & Pomphrey, N. *Rev. Geophys.* 24, 493–536 (1986).
7. Gregg, M. C. *J. Geophys. Res.* 94, 9686–9698 (1989).
8. Kunze, E., Rosenfeld, G. S., Carter, G. S. & Gregg, M. C. *J. Phys. Oceanogr.* 32, 1890–1913 (2002).
9. Winters, K. B. & D'Asaro, E. A. *J. Phys. Oceanogr.* 27, 1937–1945 (1997).
10. www.argo.ucsd.edu
11. Finnigan, T. D., Luther, D. S. & Lukas, R. *J. Phys. Oceanogr.* 32, 2988–3002 (2002).
12. Marshall, J., Jones, H., Karsten, R. & Wardle, R. *J. Phys. Oceanogr.* 32, 26–38 (2002).

Microbiology

Gut defence

Tomas Ganz

The severity of salmonella infections depends in part on how effectively the invaders are destroyed. Incisive experiments now show that host defence in the intestine centres on the aptly named defensins.

Cells that are engaged in the antimicrobial defence of mammals and birds produce defensins, a family of structurally related antimicrobial molecules^{1,2}. The production of similar small proteins is also often induced by infections in invertebrates and plants^{3,4}. Because of their abundance in infected tissues and their ability to kill a variety of microbes under laboratory conditions, defensins are thought to function as natural antibiotics. Until now, however, the evidence for their contribution to antimicrobial defence in living mammals was entirely circumstantial. That changes with the appearance of the paper by Salzman *et al.* on page 522 of this issue⁵. The authors report that they performed a genetic transplant of human defensin 5 (HD-5) into mice and observed a dramatic improvement in the resistance of the mice to intestinal infection with *Salmonella typhimurium*.

Salmonella bacteria cause several types of disease in humans and animals⁶. A diarrhoeal illness contracted by eating contaminated food is the most common consequence of *S. typhimurium* infection in humans. In this disease, the bacteria are confined to the lumen and absorptive surface of the small intestine, where they cause inflammation but do not spread through the blood to other organs. However, when *S. typhimurium* infects mice, or the related bacterium *S. typhi* infects humans, they cause typhoid fever, a much more serious illness, in which the bacteria spread from the intestine to other organs. The severity of the

disease depends to a large extent on the ability of the infected host to restrict the multiplication of the bacteria and to prevent them from penetrating the intestinal wall.

The intestinal tract is normally inhabited by a variety of resident bacteria growing at low density in the small intestine but abundantly in the colon. Paneth cells⁷ located in crypts, tiny pits throughout the small intestine (Fig. 1), release defensins and other antimicrobial substances and contribute to the ability of fluid in the small intestine to prevent the growth of invading bacteria. Previous studies had shown that *S. typhimurium* is much more sensitive to human defensin HD-5 than to the defensins produced by mouse Paneth cells.

Salzman *et al.*⁵ reasoned that if defensins function as natural antibiotics, then transgenic mice constructed to make HD-5 in their Paneth cells (HD-5 mice) should have increased resistance to infection with *S. typhimurium*. They orally infected normal and HD-5 mice with these bacteria, and observed that all HD-5 mice survived infection with doses of bacteria that killed all the normal mice within 2 days. Moreover, by 24 hours after infection, the bacterial counts in the faeces or intestines of HD-5 mice were a thousand times lower than those in normal mice, and the spread to other organs was also greatly decreased. The protective effect of HD-5 was detectable very early, with obvious differences between the HD-5 and normal mice only 6–12 hours after infection. The site of HD-5 activity was clearly in the intestine,

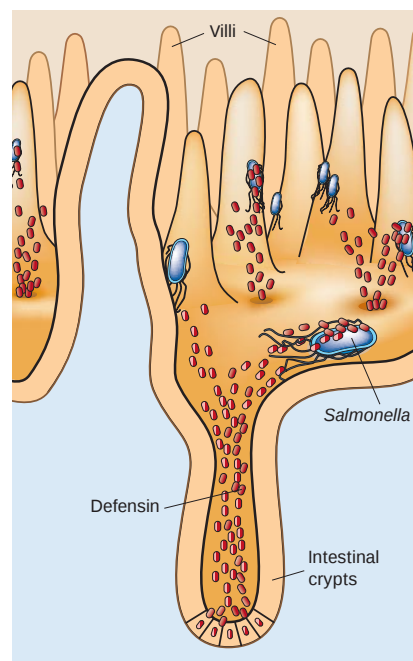


Figure 1 Bacterial invasion and the gut response. *Salmonella* bacteria attach to villi of the intestine but are attacked by defensins secreted from Paneth cells located in the intestinal crypts. Defensins are stored in large cellular granules, and are activated during or after release by the removal of a 'propeptide' by trypsin, an enzyme released by Paneth cells. In the new work, Salzman *et al.*⁵ provide solid experimental evidence that defensins act as species-specific antibiotics *in vivo*. (Redrawn from artwork by Dave Schumick, Cleveland Clinic Foundation.)

because there was no difference in the survival of HD-5 mice and normal mice when the bacteria were injected into the abdominal cavity, bypassing the intestinal secretions.

Finally, the authors also documented that the concentrations of human defensin in HD-5 mice were comparable to the concentrations of native mouse defensins, so that the observed effects were not due to unrealistically high levels of human defensins in the transgenic animals. The presence of comparable amounts of normal mouse Paneth-cell defensins, and the rapid time course of the effect of HD-5, also suggest that HD-5 acts as an antibiotic and not by indirect mechanisms such as increased recruitment of host defence cells to the infection area or priming of other immunological responses. More detailed examination of HD-5's mechanism of action in this setting should clarify this point.

Several previous studies support the idea that defensins and other antimicrobial peptides contribute to antibacterial defence in mammals. Mice that lack matrilysin, an enzyme that activates several mouse intestinal defensins, also show impaired resistance to intestinal infections⁸. However, matrilysin

deficiency is not strictly equivalent to defensin deficiency because, as well as activating defensins, matrilysin could have other functions in host defence. Mice constructed to lack another non-defensin antimicrobial peptide, CRAMP (cathelin-related antimicrobial peptide), are abnormally susceptible to skin infections with group A streptococci⁹. But CRAMP and defensins are structurally very different, so it could not be assumed that their biological roles are the same. The evidence presented by Salzman *et al.* is the strongest yet that defensins restrict microbial growth *in vivo*.

Defensins are not only made in the intestine but are abundant in other human cells and tissues. Four α -defensins closely related to HD-5 are found in human neutrophils, blood cells that specialize in the killing of bacteria and fungi in infected tissue. At least three β -defensins, which belong to a related branch of the defensin family, are secreted into the urine from the kidneys, made in infected skin, or found in other secretions that cover exposed body surfaces. Although it is likely that their functions are similar to that of HD-5, detailed studies of their specific contributions to host defence would be worthwhile. Larger human proteins, such as lysozyme and secretory phospholipase A₂, may also act as natural antibiotics in the small intestine¹⁰ and elsewhere¹¹.

It has long been known that the same bacteria can cause severe, mild or no disease, depending on the mammalian species infected¹², but the reasons for these differences have remained a mystery. The study of Salzman *et al.*⁵ suggests that humans infected

with *S. typhimurium* develop a much milder disease than do mice because human intestinal defensins are more effective against these bacteria than are their mouse counterparts. Defensins are a rapidly evolving family of peptides, and even closely related animal species often differ in the tissues in which the defensins are made and which microbes they can kill or inhibit. *Salmonella typhimurium* (Fig. 2) is one of the most common agents of bacterial diarrhoea in humans⁶, and human intestinal defensins may have evolved to be particularly effective against these bacteria. At the other extreme, humans often suffer from infections that spare other animals. By showing that a human-derived antibiotic can ameliorate an infection in mice, the new study raises the prospect that peptide antibiotics from other animal species may soon be used to help treat infections in humans. ■

Tomas Ganz is in the Department of Medicine, CHS 37-055, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California 90095-1690, USA.

e-mail: tganz@mednet.ucla.edu

1. Lehrer, R. I., Lichtenstein, A. K. & Ganz, T. *Annu. Rev. Immunol.* **11**, 105–128 (1993).
2. Lehrer, R. I. & Ganz, T. *Curr. Opin. Immunol.* **14**, 96–102 (2002).
3. Bulet, P., Hetru, C., Dimarcq, J. L. & Hoffmann, D. *Dev. Comp. Immunol.* **23**, 329–344 (1999).
4. Thomma, B. P., Cammue, B. P. & Thevissen, K. *Planta* **216**, 193–202 (2002).
5. Salzman, N. H., Ghosh, D., Huttner, K. M., Paterson, Y. & Bevis, C. L. *Nature* **422**, 522–526 (2003).
6. Zhang, S. *et al. Infect. Immunol.* **71**, 1–12 (2003).
7. Ayabe, T. *et al. Nature Immunol.* **1**, 113–118 (2000).
8. Wilson, C. L. *et al. Science* **286**, 113–117 (1999).
9. Nizet, V. *et al. Nature* **414**, 454–457 (2001).
10. Qu, X. D., Lloyd, K. C., Walsh, J. H. & Lehrer, R. I. *Infect. Immunol.* **64**, 5161–5165 (1996).
11. Ganz, T. *J. Clin. Invest.* **109**, 693–697 (2002).
12. Metchnikoff, E. *Immunity in Infective Diseases* (Cambridge Univ. Press, 1905).

Planetary science

Of asteroids and onions

John A. Wood

A refined analysis of a group of meteorites has resolved controversy over the likely structure of their parent asteroid. A layered structure is revealed, rather like an onion, whose outer regions cooled fastest.

The first stages of planetary accretion in the early Solar System produced small bodies, or 'planetesimals', of rock and ice. Most of these first-born were swallowed up by the present-day planets as they grew, but some survived as asteroids, orbiting between Mars and Jupiter where no large planet formed to consume them. These asteroids suffer continuing collisions, and rocks from their interiors may find their way to the Earth as meteorites. From such meteorites the temperature history of their parent planetesimal can be traced — this is known as thermochronometry. On page 502 of this issue, Trieloff and colleagues¹ give a comprehensive account of the temperature history of one particular planetesimal, as read from the properties of meteorites that once resided at various depths inside it.

It is known that planetesimals were heated to very high temperatures soon after they formed, almost certainly by the decay of the short-lived aluminium isotope, ²⁶Al (with a half-life of 0.7 million years), that was present in the early Solar System. The interiors of some planetesimals melted; others did not get quite hot enough to melt, but were severely metamorphosed. There are several tools that can be used to trace how the internal temperature varied over time. The peak temperature that was reached (about 1,100 K in the asteroid studied by Trieloff *et al.*¹) can be estimated from the relative abundance of ferrous iron in two crystal forms (orthorhombic and monoclinic) of

the silicate minerals known as pyroxenes². The decay scheme of uranium into lead can be exploited to work out the time that has elapsed since the rocks cooled below about 720 K, because thereafter, lead isotopes created by uranium decay were not able to diffuse away from the sites of their uranium-atom parents³. Similarly, potassium-to-argon decay tells us how long it has been since the rocks cooled through about 550 K, into a temperature range where argon atoms created by potassium decay could not diffuse away from the sites where the decay occurred⁴.

A particularly ingenious thermochronometric tool — and one used to great effect by Trieloff *et al.*¹ — relies on the fact that the (relatively) short-lived plutonium isotope, ²⁴⁴Pu (with a half-life of 80 million years), was present in the newly formed planetesimals. Plutonium, which decays by spontaneous fission, creates radiation damage in the crystals where it is sited. At high temperatures, the trails of lattice damage created by fission fragments 'heal' (anneal). But at low temperatures they persist, and chemical etching of crystals containing trails makes the fission tracks visible through a light microscope. The density of tracks can be measured and related to the amount of ²⁴⁴Pu remaining in the system when annealing became impossible. This density can therefore also be related to time. Plutonium in meteorites tended to concentrate in phosphate minerals, and below about 390 K those

Figure 2 *Salmonella typhimurium*, shown here in culture. These bacteria usually have much more severe effects in mice than in humans.



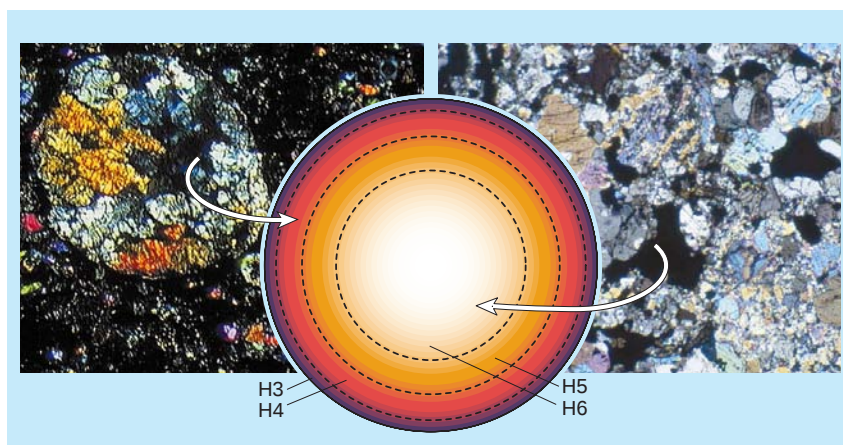


Figure 1 The onion-shell model of an asteroid interior. A suite of meteorites known as the H-chondrites are believed to have originated from a single asteroid. Trierloff *et al.*¹ have quantified the cooling histories of these meteorites and confirmed the validity of the onion-shell model: the H6 chondrites that show evidence of the most severe metamorphism were the most deeply buried; H5, H4 and H3 chondrites, which were less affected by thermal metamorphism, came from shallower levels. The different textures of the chondrites are seen in the micrographs of thin sections of H4 (left) and H6 (right) chondrites.

minerals were not able to heal radiation damage. The minimum annealing temperature for orthorhombic pyroxene is higher, about 550 K (although the pyroxene did not contain plutonium, tracks were created in it where it butted against phosphate minerals). As some ²⁴⁴Pu decay occurred between the time when a meteorite cooled to 550 K and when it cooled to 390 K, the fission track density is smaller in phosphate minerals than in pyroxene. This difference in track density means that the cooling time between 550 K and 390 K can be calculated⁵.

Trierloff *et al.*¹ studied a chemically distinct suite of meteorites, known as H-group chondrites, that are believed to derive from the same asteroid. The samples they studied show different degrees of thermal metamorphism, which are denoted by their classification, ranging from H6 (most metamorphosed) to H3 (least so)⁶. The planetesimal model that accounts most straightforwardly for these degrees of metamorphism would have the H6 chondrites near its centre, where an internal heat source would heat the rock most and where it would stay hot the longest. At progressively shallower depths the H5, H4 and H3 rock was heated less and cooling occurred more quickly⁷. This is the 'onion-shell' model of meteorite parent bodies (Fig. 1).

Evidence that chemically distinctive subsets of meteorites are derived from onion-shell parent bodies has been sought for decades without success. Thermochronometric analysis produced inconsistent, conflicting results, calling into question the validity of the onion-shell model. Perhaps, instead, a collision between asteroids caused the H-chondrite parent body to break up while it was still hot, and thereafter the fragments had diverse cooling histories⁸. Maybe there was more than one H-chondrite

parent body, and the suite of H-chondrites in fact represents samples of several planetesimals, each with a different thermal history⁹. Or could the heat source have been external rather than internal¹⁰?

Trierloff *et al.* now dispel these doubts by showing that the thermochronometric data for H6, H5 and H4 chondrites can be fitted into a straightforward model of a planetesimal with a radius of about 100 kilometres and an onion-shell structure, which was internally heated by ²⁶Al decay and cooled over about 100 million years. (H3 chondrites are too fine-grained and poorly equilibrated to yield useful data.) This advance was made possible

by three factors. First, the authors carefully selected the samples that they studied, excluding any chondrites that showed mineralogical evidence of shock — the damage produced in some meteorites by asteroid collisions, which can confuse the thermochronometric record. Second, they included high-quality measurements of potassium–argon ages in their analysis. Third, the fission-track technique was refined by finding a way to remove extraneous tracks caused by cosmic rays, and by correcting for differences in the track-registration efficiencies between phosphate and pyroxene minerals.

This paper represents a posthumous triumph for co-author Paul Pellas, a highly respected and charismatic physicist at the Muséum National d'Histoire Naturelle in Paris, whose principal life's work was the development of the fission-track technique of thermochronometry.

John A. Wood is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.
e-mail: jwood@cfa.harvard.edu

1. Trierloff, M. *et al.* *Nature* **422**, 502–506 (2003).
2. Bunch, T. E. & Olsen, E. *Contrib. Mineral. Petrol.* **43**, 83–90 (1974).
3. Göpel, C., Manhès, G. & Allègre, C. J. *Earth Planet. Sci. Lett.* **121**, 153–171 (1994).
4. Turner, G., Enright, M. C. & Cadogan, P. H. *Proc. Lunar Planet. Sci. Conf.* **9**, 989–1025 (1978).
5. Pellas, P., Fieni, C., Trierloff, M. & Jessberger, E. K. *Geochim. Cosmochim. Acta* **61**, 3477–3501 (1997).
6. Van Schmus, W. R. & Wood, J. A. *Geochim. Cosmochim. Acta* **31**, 747–765 (1967).
7. Miyamoto, M., Fujii, N. & Takeda, H. *Proc. Lunar Planet. Sci. Conf.* **12B**, 1145–1152 (1981).
8. Grimm, R. F. J. *Geophys. Res.* **90** (B2), 2022–2028 (1985).
9. Scott, E. R. D. & Rajan, R. S. *Geochim. Cosmochim. Acta* **45**, 53–67 (1981).
10. Sonett, C. P., Colburn, D. S. & Schwartz, K. *Nature* **219**, 924–926 (1968).

Cell motility

Making streams

Richard H. Kessin

Cells of the slime mould *Dictyostelium discoideum* can move in a kind of close-order drill. New evidence suggests that they do this by secreting a chemical attractant specifically from their rear, luring the cells that follow.

In 1944, the microbiologist Ernest Runyon put amoebae of the then little studied slime mould *Dictyostelium discoideum* on both sides of a piece of porous cellophane¹. Over the next few hours, the amoebae on one side started to gather at a central point, moving at first individually and then, tail to head, in streams. Astonishingly, so did the amoebae on the other side of the cellophane, forming mirror-image streams. Runyon presumed that this was due to a molecule, secreted through the cellophane, that attracted cells on the other side. This turned out to be the first observation that a diffusible attractant molecule is involved in

guiding aggregation. It would be more than two decades before John Bonner and his colleagues² discovered that this molecule was cyclic AMP, secreted by the migrating amoebae. Carole Parent and colleagues have continued with this line of research, and now show in *Cell*³ that the *D. discoideum* enzyme that produces cAMP is located specifically at the rear of the migrating cells — perhaps explaining how they move in a stream.

Many different cell types undergo chemotaxis — movement up a gradient of an attractive chemical towards its central source. The ability of *D. discoideum* amoebae to do this, and consequently to

form aggregates (Fig. 1), is crucial to their survival in harsh environmental conditions. Failure to enter the aggregate, which later differentiates into an elaborate fruiting body producing tough, resistant spores, is a distinct disadvantage to a cell. Amoebae cannot form spores alone: they move or they die. There might also be particular advantages to marching in close-order files, such as the ability to lay down extracellular matrix as a group. This matrix, called the slime sheath, appears at about the time of stream formation and could help the cells to move over rugged surfaces — *D. discoideum* amoebae can clamber efficiently over dirt and most other substrates.

Since Bonner's discovery that cAMP is needed for streaming in *D. discoideum*, many of the elements and mechanisms involved have been worked out. (In experiments, cAMP is generally used as the central source of chemoattractant, as well as being released naturally by the migrating amoebae.) We now know that this organism has at least four types of cAMP receptor, as well as several unusual molecules that regulate adenylyl cyclase — the enzyme that synthesizes cAMP from ATP (reviewed in refs 4–6). An extracellular signal of cAMP is detected by the receptors, which are distributed over the entire periphery of the cell. Specific lipids then accumulate in the membrane at the front edge of the cell, and recruit certain molecules, including CRAC (for 'cytosolic regulator of adenylyl cyclase') and AKT/PKB, an enzyme involved in organizing cellular movement^{7,8}. This series of events ensures that the cell moves towards the source of cAMP, and also leads to the activation of adenylyl cyclase and the further production

of cAMP, a large proportion of which is duly secreted. And so the process continues in the next cell in line.

The expectation has always been that adenylyl cyclase itself would be located everywhere along the cell membrane, much like the cAMP receptors. But that turns out not to be the case. Parent and colleagues³ have found that during chemotaxis, streaming *D. discoideum* amoebae localize adenylyl cyclase at their rear (in their 'uropod'), putting the enzyme in close contact with the leading edge of the following cell, but quite far from the receptors that first detect cAMP. Deleting the adenylyl cyclase still permits chemotaxis towards a central — natural or artificial — source of cAMP, but not the extensive head-to-tail line-up of cells that is normally seen. The authors reasoned that if, as these results suggest, the position of adenylyl cyclase in the rear is essential for creating streams, then mixing cells that lack adenylyl cyclase with wild-type cells should terminate the chains. That proved to be the case (Fig. 2). It has been known for decades⁹ that amoebae enter a stream behind a cell that is already in the queue. This behaviour could be explained if the moving cells secreted cAMP specifically from their uropods, as suggested by the location of adenylyl cyclase — this remains to be shown definitively.

Parent and colleagues also found that the localization of adenylyl cyclase to the uropod requires that the cell be polarized. This happens gradually as *D. discoideum* starts to starve in harsh conditions — a prelude to migration into an aggregate. Moreover, localization does not depend on protein kinase A, which is the main target of cAMP within the cell and a central regulator of events during *Dictyostelium* development, but it does require an intact internal 'skeleton' made up of the actin protein. Membrane-bounded intracellular sacks (vesicles) containing adenylyl cyclase also seem to be needed. The authors showed that a permanently active form of adenylyl cyclase did not become as enriched as normal in the uropod, and so, as predicted, amoebae that had this mutant enzyme did not form streams during chemotaxis. Wild-type cells could insert adenylyl cyclase in the uropod on chemotactic stimulation. But in mutant cells the enzyme seemed to be trapped in vesicles that did not reach their proper destination.

All of this leaves open the question of how events at the leading edge of a migrating cell — increasingly well worked out by the groups of Parent, Devreotes, Firtel and others — lead to the activation of adenylyl cyclase in the rear. Whatever the answer, Parent and colleagues' data³ provide a mechanism by which cells can increase the efficiency of chemotaxis. Presumably, the close contact between the adenylyl cyclase in the rear of one cell and the front of the cell behind provides some of the advantages of

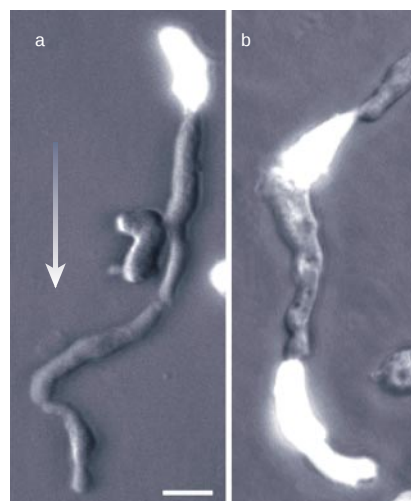


Figure 2 The importance of proper molecular positioning to streaming in *Dictyostelium discoideum*. Parent and colleagues³ have found that the enzyme adenylyl cyclase must be located at the rear of migrating *D. discoideum* amoebae for streaming to occur. Presumably, this localization enables the attractant chemical cAMP, produced by adenylyl cyclase, to be released specifically from the rear and thereby recruit another cell directly to it. a, Cells with mutant adenylyl cyclase were labelled with green fluorescent protein (GFP), and appear white here. They were then mixed with unlabelled wild-type cells and allowed to form streams. Usually, the mutant cells terminate a stream. b, Wild-type cells labelled with GFP were allowed to form a stream with other wild-type cells. Wild-type cells can enter a stream and attract other cells. The large arrow indicates the direction of migration. Scale bar, 10 μ m. (Reproduced from ref. 3 with permission from Elsevier.)

a neurological synapse, restricting the signal and keeping the amoebae heading in the right direction as quickly as possible. Perhaps mammalian cells can do the same, whether they are white blood cells moving to sites of inflammation, skin cells participating in wound healing, or embryonic cells during the heroic voyages of development. ■

Richard H. Kessin is in the Department of Anatomy and Cell Biology, Columbia University, 630 West 168th Street, New York, New York 10032, USA. e-mail: rhks@columbia.edu

- Runyon, E. H. in *The Collecting Net* 88 (Marine Biological Lab., Woods Hole, Massachusetts, 1942).
- Konijn, T. M., van de Meene, J. G. C., Chang, Y. Y., Barkley, D. S. & Bonner, J. T. *J. Bacteriol.* 99, 510–512 (1969).
- Kriebel, P. W., Barr, V. A. & Parent, C. A. *Cell* 112, 549–560 (2003).
- Parent, C. A. & Devreotes, P. N. *Science* 284, 765–770 (1999).
- Chung, C. Y., Funamoto, S. & Firtel, R. A. *Trends Biochem. Sci.* 26, 557–566 (2001).
- Kessin, R. H. *Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity* (Cambridge Univ. Press, 2001).
- Insall, R. et al. *J. Cell Biol.* 126, 1537–1545 (1994).
- Chung, C. Y., Potikyan, G. & Firtel, R. A. *Mol. Cell* 7, 937–947 (2001).
- Shaffer, B. M. in *Primitive Motile Systems in Cell Biology* (eds Allen, R. D. & Kamiya, N.) 387–405 (Academic, New York, 1964).

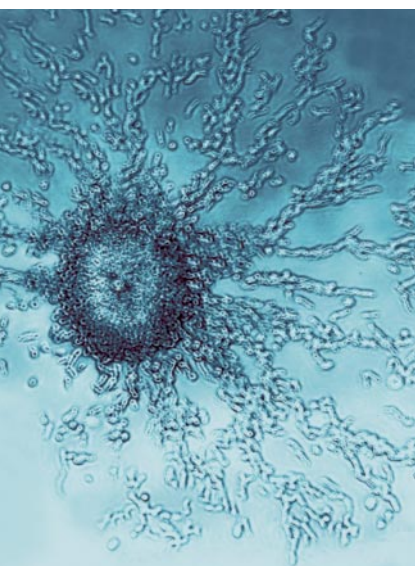


Figure 1 A gathering of *Dictyostelium discoideum* amoebae. Cells can be seen migrating — some individually, and some in streams — towards a central point. Aggregation territories can be as much as a centimetre in diameter.

Behavioural ecology

Coots count

Malte Andersson

American coots distinguish their own eggs from the eggs that other female coots lay in the same nest. They use a variety of tactics to minimize the adverse reproductive effects of the parasitic behaviour.

Cuckoos are infamous for laying their eggs in the nests of other bird species: the unfortunate hosts raise the parasitic chicks, reducing or even nullifying their own reproductive output. It is less well known that 'brood parasitism' among females of the same species is common in certain fishes, insects and birds¹. In this conspecific brood parasitism, where females parasitize others of their own species, the host parents will bear a high cost when they feed chicks that are not their own.

Ways of reducing these costs are therefore expected to be favoured by natural selection. On page 495 of this issue, Bruce Lyon² describes how an aquatic bird, the American coot *Fulica americana* (Fig. 1), uses several remarkable behavioural tactics to this end. In this species, a parasitic chick that survives usually does so at the cost of a host chick. But the host parents can often identify the eggs of parasites, and reject or otherwise disadvantage many of them.

With some exceptions^{3–6}, discrimination against parasitic eggs seems to be a rare defence in conspecific brood parasitism. In many species, hosts may not be able to tell the difference between their own and alien eggs¹. Coot eggs vary considerably in colour, however, and Lyon finds that coots can not only often distinguish their own eggs from those of a parasite, but — remarkably — can count them, disregarding the number of parasitic eggs in the clutch. This surprising cognitive feat makes good adaptive sense. Like many other birds, coots are indeterminate layers⁷; that is, they regulate egg production using external stimuli, such as the number or surface area of the eggs already in the nest, to control egg development and final clutch size. If the parent cannot distinguish between its own and parasitic eggs, the latter can cause it to stop egg production too early, reducing the final clutch size below that necessary for its own maximum reproductive success.

As with most birds' eggs, there is little difference in surface texture between eggs laid by different female coots. So although responses to a tactile stimulus cannot be completely ruled out, it seems unlikely that this is the distinguishing cue. By contrast, to the human eye, the coloration of coot eggs — a conspicuous pattern of dark spots on a brown background — varies considerably among females (Fig. 1, inset). Visual cues therefore seem to offer ample scope for egg discrimination. Lyon² indeed found that the visual

differences between eggs of host and parasite were larger in nests where parasitic eggs were rejected than where they were accepted. For the host to be able to reject a parasitic egg, therefore, it seems that there must be a large enough difference in the eggs' appearance.

Lyon compared coot pairs that accepted parasitic eggs ('acceptors') with pairs that rejected them ('rejectors'). He found that acceptor females produced a lower final number of their own eggs. Rejector females, however, did not reduce their own final clutch size. Evidently they recognized and counted their own eggs, avoiding a maladaptive reduction in the number of their own eggs in response to the added parasite eggs. Later on, rejector hosts discarded parasitic eggs by burying them under nest material, or moved them to peripheral positions in the clutch. The consequence was to prevent or delay the hatching of parasitic chicks, so reducing competition with the host's true offspring. On average, such behaviour halved the effect of parasitism on the reproductive success of the hosts.

Some aspects of this defensive behaviour raise more questions. Why do coots eject

cracked or rotten eggs from the nest, but bury parasitic eggs? And why do they move some parasitic eggs to the periphery of the clutch rather than reject them altogether? Lyon suggests that the latter behaviour may represent an intermediate form of defence, used when the bird is uncertain whether the egg is parasitic.

From an ecological perspective, Lyon's results are striking in what they tell us about the costs of brood parasitism, and defences against it. They also shed light on cognition and counting by animals, in a natural context where reproductive success is at stake. There is suggestive evidence⁸ that counting may be important in such circumstances. But it is rarely as clear as here.

Usually, the number of items counted in natural situations is small, as is the case for coots counting their eggs. With proper experimental training, however, pigeons and rats have made surprisingly accurate choices in complex tasks that involve counting to up to several dozen⁹. Many animals apparently have a brain wiring that in the right circumstances can support competent counting without verbal symbolic representation of numbers. Lyon's findings provide a fascinating example of how this capacity is put to good use in the wild. Who would have expected that from a birdbrain?

There are plenty of open questions. For example, what is the genetic basis and heritability of the visual appearance of eggs, and does this have consequences for parasitism involving relatives? Does parasitism create



Figure 1 Coots in a flap. These birds fight over territorial borders, territory size setting the limit for the number of chicks that can be successfully fed by the parents. Hence the evolution of brood parasitism to bypass this constraint. The inset shows a parasitized nest — the two darker eggs are the ones laid by a brood parasite.

BRUCE LYON

selection for increased variability in egg coloration, making it easier to detect parasitic eggs⁵? A combination of Lyon's incisive field techniques with genetics, and with molecular determination of parasitism and parentage^{10,11}, seems likely to provide further insights into the cognitive and tactical aspects of brood parasitism and reproductive behaviour. ■

Malte Andersson is in the Department of Zoology, University of Gothenburg, Box 463, SE 405 30 Göteborg, Sweden.
e-mail: malte.andersson@zool.gu.se

Nuclear physics

Into the fission valley

Peter Möller and Arnold J. Sierk

Why are new elements difficult to make? Fusion of two nuclei to produce heavy elements seems to be hindered by a competing process of 'quasi-fission'. New work builds a more complete picture.

On Earth, elements heavier than uranium do not exist in easily measurable quantities, because they become increasingly unstable against radioactive decay. Some heavier elements can be artificially created through the collision and fusion of other nuclei, and over the past 60 years about 20 new elements have been added to the periodic table — up to an atomic number of at least 110. But fusion becomes less successful when the projectile and target nuclei are chosen to form a product with nucleon number, A , greater than 220 (more than 90 protons and 130 neutrons). For example, in some experiments¹ only about one out of 10^{18} nuclei incident on a target leads to the creation and detection of the desired new element.

In *Physical Review Letters*, Hinde, Dasgupta and Mukherjee² present a detailed analysis of this inhibition of fusion near $A = 220$. They made a careful comparison of fusion cross-sections (or probabilities) from their own experiment on the reaction $^{16}\text{O} + ^{204}\text{Pb}$ with other experiments on $^{40}\text{Ar} + ^{180}\text{Hf}$, $^{48}\text{Ca} + ^{172}\text{Yb}$, $^{82}\text{Se} + ^{138}\text{Ba}$ and $^{124}\text{Sn} + ^{96}\text{Zr}$, all of which lead to the same compound system, ^{220}Th . The results show that it is much more difficult to make ^{220}Th with more symmetric combinations of target and projectile than with the most asymmetric combination ($^{16}\text{O} + ^{204}\text{Pb}$) — specifically about ten times more difficult. Hinde *et al.* propose that this is due to competition with the process of 'quasi-fission'. Colliding nuclei form a composite at the onset of the fusion process, but the composite may break up, or undergo fission, before fusion is complete. True fission occurs after the formation of an equilibrated compound nucleus; quasi-fission results from the much faster breakup of a partially fused composite.

Today's theories of heavy-ion collisions are mainly macroscopic; the energy of the

1. Davies, N. B. *Cuckoos, Cowbirds and Other Cheats* (Poyser, London, 2000).
2. Lyon, B. E. *Nature* **422**, 495–499 (2003).
3. Arnold, T. W. *Condor* **89**, 675–676 (1987).
4. Bertram, B. C. R. *The Osprey Communal Nesting System* (Princeton Univ. Press, 1992).
5. Jackson, W. M. in *Parasitic Birds and Their Hosts* (eds Rothstein, S. I. & Robinson, S. K.) 406–416 (Oxford Univ. Press, 1998).
6. Jamieson, I. G., McRae, S. B., Simmons, R. E. & Trewby, M. *Auk* **117**, 250–255 (2000).
7. Haywood, S. Q. *Rev. Biol.* **68**, 33–60 (1993).
8. Seibt, U. *Behav. Brain Sci.* **11**, 597–598 (1988).
9. Hauser, M. D. *Am. Sci.* **88**, 144–151 (2000).
10. Andersson, M. & Åhlund, M. *Ecology* **82**, 1433–1442 (2001).
11. Birkhead, T. R., Burke, T., Zann, R., Hunter, F. M. & Krupa, A. P. *Behav. Ecol. Sociobiol.* **27**, 315–324 (1990).

colliding system can be written as a sum of the repulsive electrostatic energy and the attractive nuclear energy. After the colliding nuclei touch, these energies must be calculated for the combined system as functions of its shape. The resulting 'energy landscape' (a multi-dimensional potential-energy surface in deformation space) has a strong influence on the dynamics of the fusion process.

In macroscopic models, the energy landscape changes relatively slowly with proton and neutron number and with deformation of the shape of the nucleus away from a simple sphere. But microscopic effects, which arise because the protons and neutrons in the nuclei obey quantum-mechanical laws, vary much more rapidly as the neutron and proton numbers and the shape change, sometimes producing large differences between the behaviour of systems with only slightly different nucleon number. Microscopic effects are not often included in theoretical studies of nuclear collisions, but we believe that they should be considered more carefully. There are other factors, too: how dissipation converts the kinetic energy of the projectile into internal excitation energy of the fusing system; and the effect of the relative orientation of target and projectile if one or both of them are deformed (around 50% of stable nuclei are not spherical).

In their paper, Hinde *et al.*² consider various explanations for the inhibition of fusion seen in the data. First, they discuss the idea of the 'extra push' — a colourful misnomer used to describe a dynamical threshold. For heavy compound systems, extra kinetic energy (more than is needed to bring the nuclei into contact after overcoming their electrostatic repulsion) is required for them actually to fuse and form a single nucleus. A good analogy is a skier crossing a mountain range, starting out with some initial energy. For



100 YEARS AGO

Concerning the recently discovered heat emission from radium, it is perhaps worth noting that it appears to be connected with, and is probably an immediate consequence of, the remarkable observation by Rutherford that radium emits massive positively-charged particles, which are probably atoms, with a velocity comparable to one-tenth of the speed of light... Because it is easy to reckon that the emission of a million heavy atoms per second, which is a small quantity barely weighable in a moderate time such as a few weeks (being about the twentieth part of a milligramme per century), with a speed equal to one-tenth that of light, would represent an amount of energy equal to one thousand ergs per second; that is to say, would correspond to heat enough to melt a milligramme of ice every hour. And inasmuch as these atoms are not at all of a penetrating kind, but are easily stopped by obstacles, they would most of them be stopped by a small thickness of air, and their energy would be thus chiefly expended in the immediate proximity of the source, which source would thereby tend to be kept warm.

From *Nature* 2 April 1903.

50 YEARS AGO

Before the War one could make materials artificially radioactive by bombardment in big machines like cyclotrons, or by using relatively weak neutron sources. In the cyclotron, one can generally only use one target at a time and the irradiation is therefore costly. The weak neutron sources induce only weak activities. Therefore only a few research workers profited from the radioisotopes which one could produce in these ways. The situation changed suddenly with the discovery of fission of uranium in 1939. This discovery showed that chain reactions were possible in which more neutrons are created than used. Accelerated by war research, the first chain-reacting atomic pile was working on December 2, 1942, in Chicago... New radioisotopes were quickly discovered and the chart of radioactive isotopes started to expand. To-day, there are more than six hundred radioactive isotopes, of which, however, only some hundred can be made conveniently in an atomic pile. For most elements there is at least one usable radioactive isotope. The only notable exceptions are the two elements nitrogen and oxygen for which no convenient radioactive isotope exists.

From *Nature* 4 April 1953.

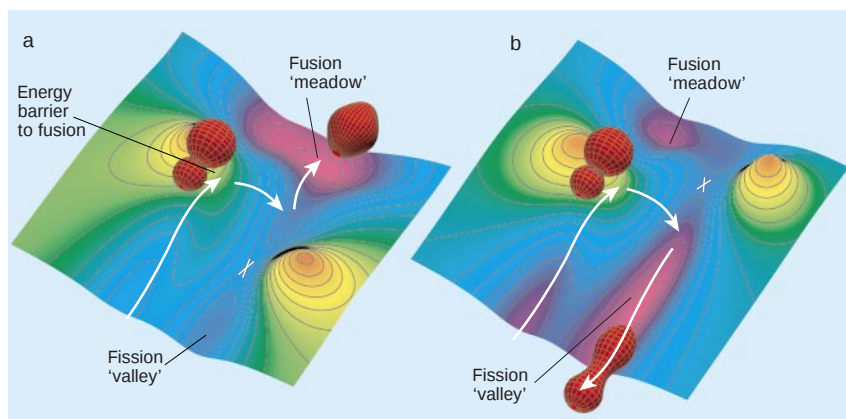


Figure 1 Schematic energy landscapes for fusing nuclei. **a**, For lighter nuclei, the pair move up the 'fusion valley', opposed by electrostatic repulsion. The maximum in the potential along the initial path ('fusion barrier') occurs near the point where they first touch. If there is only enough initial energy to reach this point, the forces push the system to the right, leading inside the fission saddle point (indicated by a cross). Then the system falls towards the fused ground state (pink area). **b**, For a heavier system, the nuclei pass outside the saddle point, eventually falling down the fission valley — unless they are given sufficient extra initial energy to drive them towards fusion.

lighter systems (Fig. 1a), the process of fusion corresponds to climbing up to the head of a fusion valley (overcoming electric repulsion) and surmounting a barrier (where the short-range nuclear attraction balances the electric repulsion) to access a meadow beyond (the lower energy of the fused nuclei). After passing over this barrier, the nuclei come into contact, and the system, or skier in our analogy, is on the side of a hill, which is due to the forces trying to deform the nuclei. Once he has surmounted the barrier peak, even with no residual energy, the skier would naturally descend into the depression beyond.

For heavy systems (Fig. 1b), in which the electric forces are relatively stronger, the topography is different. For one thing, the 'fusion meadow' becomes much smaller, and has higher energy. The path to the smaller meadow lies along the side of a slightly steeper hill. So if the skier started up the valley with just enough energy to reach the peak of the barrier, he would find himself pushed sideways, moving away from the meadow down a separate 'fission valley' below his initial path on the side of the hill.

By starting out with additional energy, the skier would arrive at the touching point with residual forward momentum, and could continue along the side of the hill far enough to drop into the meadow. So even in the absence of friction, extra energy above the barrier peak (dynamical threshold energy) is needed for heavy systems to fuse. If there are dissipative processes (such as energy coupling from the motion of the nuclei into internal excitations when they come into close proximity), even more energy is needed for fusion. Calculations in macroscopic models show that a fairly rapid transition occurs between the two situations sketched above, owing to the shift of the fission saddle point, when the target and

projectile masses add up to $A \approx 220$. This is why it is relatively easy to make lighter nuclei through fusion and more difficult to make heavier nuclei. Experiments and macroscopic theory seem to agree well on this feature of fusion reactions.

But there are other details to be taken into account for heavy systems. Hinde *et al.*² also looked at the effect of the asymmetry of a system on the fusion probability. In very asymmetric systems, the larger body tends to absorb the smaller one, but in more balanced systems there tends to be a transfer of mass from the heavy to the light partner. We would caution against invoking this macroscopic mechanism too strongly, however, as microscopic effects may dominate the gentler macroscopic forces.

Another consideration is that in the quantum world a reaction is more successful if the initial state highly resembles the final state. The final state in this case is a single spherical, or almost spherical, nucleus. A very small projectile touching a very large target more closely resembles this final state than do two nuclei of more similar size. Even in a classical context, less matter needs to be moved in the very asymmetric case.

If the target or the projectile is deformed, the relative orientation of the nuclear symmetry axes should also be considered. Hinde *et al.*² argue, as others have done³, that if an undeformed nucleus hits a (prolately) deformed one at its equator, this compact configuration might be less hindered than other collision orientations, again because of the greater compactness and similarity to the final configuration. Of course, if both target and projectile are deformed, many other intriguing possibilities arise⁴.

Microscopic effects should not be overlooked. For example, the cross-section for creating isotopes of 'darmstadtium', the

element with atomic number 110, is five times as large for $^{64}\text{Ni} + ^{208}\text{Pb}$ as for $^{62}\text{Ni} + ^{208}\text{Pb}$. Such rapid cross-section variations are not predicted in macroscopic models. In another study, Satou *et al.*⁵ saw similar differences between the two reactions $^{82}\text{Se} + ^{134}\text{Ba}$ and $^{82}\text{Se} + ^{138}\text{Ba}$ (leading to ^{216}Th and ^{220}Th , respectively): the second reaction is much more productive than the first. The isotope ^{138}Ba has extra binding energy because the outermost shell of neutrons in its nucleus is full. This extra binding (about 5 MeV) will tend to resist deformations away from the fusion path.

In our skier analogy, the full spherical shell may create a ditch along the side of the hill that the skier traverses, enabling him to follow the fusion path without needing to start with the extra momentum apparently necessary in the ^{134}Ba case. But it is also possible that, if the neutron shell is not full (as in ^{134}Ba), the nucleus would dissipate more energy because there are more low-lying energy levels that could be excited in the early stages of the interaction of the colliding nuclei. Trying to isolate the microscopic potential-energy effects from dissipative dynamics remains a problem in both theory and experiment.

Some of the more efficient reactions for the production of heavy nuclei have involved Pb targets, which also have resistance to deformation because both the proton and neutron shells are full, or almost full⁶. It is conceivable that the higher fusion probability for $^{16}\text{O} + ^{204}\text{Pb}$ is not due to the macroscopic 'swallowing-up' effect discussed above, but to the extra binding (10 MeV) of the nearly closed-shell ^{204}Pb nucleus. It is tantalizing that the next-highest cross-section in the ^{220}Th system analysed by Hinde *et al.*² involves the $^{82}\text{Se} + ^{138}\text{Ba}$ system, which has the most extra binding (5 MeV) of the other, less asymmetric channels studied.

The conclusion from the work of Hinde *et al.* is that the big picture — of valleys, mountains and meadows — is well described by some macroscopic models, but considerable effects arise because of microscopic details. The machinery for calculating these microscopic effects has developed rapidly in recent years, and may soon be used to help understand experimental data, with an eye to predicting promising new reactions to explore.

Peter Möller and Arnold J. Sierk are in the Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. e-mail: moller@moller.lanl.gov

- Hofmann, S. *et al.* *Europhys. J.* **14**, 147–157 (2002).
- Hinde, D. J., Dasgupta, M. & Mukherjee, A. *Phys. Rev. Lett.* **89**, 282701 (2002).
- Möller, P. & Iwamoto, A. *Nucl. Phys. A* **575**, 381–411 (1994).
- Iwamoto, A., Möller, P., Nix, J. R. & Sagawa, H. *Nucl. Phys. A* **596**, 329–354 (1996).
- Satou, K. *et al.* *Phys. Rev. C* **65**, 054602 (2002).
- Hofmann, S. & Münzenberg, G. *Rev. Mod. Phys.* **72**, 733–767 (2000).

Cell biology

Tumour suppressors in the wrecking yard

Proc. Natl Acad. Sci. USA **100**, 3263–3268 (2003)

The proteasome is the wrecking yard of the cell, where proteins are broken down into smaller parts. Classically, potential protein victims are earmarked by becoming linked to chains of the small protein ubiquitin. But, very rarely, different systems can kick in.

Robert F. Kalejta and Thomas Shenk now show that tumour-suppressor proteins from the retinoblastoma (Rb) family, whose job is to put the brakes on cell division, are targeted for destruction by a protein called pp71. This is encoded by human cytomegalovirus, a member of the herpes family. The proteasome is still involved, as chemicals that inhibit its function prevent the Rb proteins from being broken down. Yet how pp71 directs its victims to the proteasome is not yet known.

The results prompt speculation that cytomegalovirus has found a neat way of manipulating its cellular hosts. As Rb proteins are degraded, cells could be pushed back into the cell-division cycle and begin to proliferate, allowing the virus to replicate and spread.

Helen R. Pilcher

Ecology

Ecosystem engineering

Ecology **84**, 682–690 (2003)

Caterpillars make oak trees better environments for other insects. John T. Lill and Robert J. Marquis have found that, without one industrious species, the diversity of insect herbivores on trees falls.

***Pseudotelphusa* caterpillars build a tent by tying together a pair of adjacent leaves**



A leaf tie and (inset) a *Pseudotelphusa* caterpillar. The paler patches in the main picture are where the leaf material has been consumed by the caterpillar within.

on white oaks (*Quercus alba*). They take refuge in this shelter, and slowly eat their home from within (see picture). Lill and Marquis removed caterpillars from 93 trees, and then made artificial shelters on 62 of them — to make sure that it was this that made the difference, and not some other aspect of caterpillar biology.

On trees without caterpillars or artificial shelters there was a steep reduction in the number of species of leaf-chewing insects, including other caterpillars, sawflies and beetles. It seems that the shelters protect other species from inclement weather and predators, and many of the squatters seek them out to lay their eggs on. White oak is a dominant tree in the forests of eastern North America. So by acting as ecosystem engineers, the common shelter-building caterpillars may be having a significant effect on the environment as a whole.

John Whitfield

Nonlinear optics

Braking at room temperature

Phys. Rev. Lett. **90**, 113903 (2003)

Light slowed to a snail's pace might have uses in optical information technologies and quantum computing. But putting the brakes on the fastest thing in the Universe has previously been possible only at temperatures within a few degrees of absolute zero or within exotic environments. Matthew S. Bigelow and colleagues have now achieved it in a common solid-state material — ruby — at room temperature, bringing technological applications within sight.

'Slow light' was first reported in ultra-cold metal-atom vapours in which laser beams manipulate the refractive index of the medium. This reduces the propagation speed of a light pulse dramatically; in the extreme case the light can be stopped entirely, so that pulses can be 'stored'. A report last year of slow light in a solid metal oxide made the business look more practical — but this still required cooling to 5 K. Bigelow *et al.* use a different optical method — spectral hole burning — to control the refractive index of a ruby crystal, allowing them to slow light to about 57 m s^{-1} at room temperature. The technique is easy to implement and might find its way into optical information technology, for example to make compact, chip-scale optical delay lines.

Philip Ball

Cancer

Building blood vessels

Cancer Cell **3**, 219–231 (2003)

In order to survive, grow and spread, tumours must establish their own blood supply. Robert A. Weinberg and colleagues have found that a key step in this process is the repression of a naturally occurring

inhibitor of vessel formation, thrombospondin-1 (Tsp-1), by a cancer-causing form of the Ras protein.

The authors studied human epithelial and kidney cells expressing either high or low levels of this 'oncogenic' Ras. 'High-Ras' cells formed tumours when injected into immunodeficient mice, as did 'low-Ras' cells engineered to express low levels of Tsp-1. In contrast, when high-Ras cells were engineered to express high levels of Tsp-1, they formed smaller tumours, which struggled to establish their own blood supply.

Various other experiments led Weinberg and colleagues to conclude that vessel formation involves a complex cascade of signals, triggered by Ras, that ultimately activates a second oncogenic protein, Myc. This cascade, which is active in several human breast cancer cell lines, results in low levels of Tsp-1 and so leads to the formation of new blood vessels and tumour growth. The researchers hope that their findings will provide new targets for treatments aimed at inhibiting tumour growth.

Helen R. Pilcher

Inorganic chemistry

Molecules lend a hand

Angew. Chem. Int. Edn **42**, 1293–1296 (2003)

Nature distinguishes left from right through its failure to preserve left–right symmetry in some interactions between subatomic particles. This so-called parity violation creates a difference in energy between the two enantiomers of chiral molecules, which has been implicated in the origin of biomolecular handedness. But the difference is so tiny — a factor of about 10^{-17} for L and D amino acids — that it typically lies orders of magnitude below the detection limits of current techniques. Undeterred, Peter Schwerdtfeger and colleagues have set out to design chiral molecules that might differ sufficiently in energy to make parity violation measurable.

Because the size of the effect scales approximately with the fifth power of the nuclear charge on an atom, the researchers figured that they must target enantiomers containing heavy-metal elements. But they also wanted to find compounds that are stable and whose synthesis seems feasible. Their targets included chiral organometallic complexes of bismuth, rhenium and iridium, for each of which the researchers used quantum-chemical methods to calculate the energy differences. All the molecules have analogues synthesized previously, and their parity violation should be strong enough to be just about detectable with spectroscopic techniques.

Philip Ball

The dark side

Masataka Fukugita

Evidence for the existence of dark matter in the Universe is overwhelming. But how close are we to finding out what it really is?

The story of dark matter began in 1933. Fritz Zwicky, one of the greatest astronomers of the twentieth century, announced that the Coma cluster of galaxies ought to contain more matter than is inferred from optical observations¹. He had found that many of the thousands of galaxies in the cluster move at speeds faster than the escape velocity expected from the amount of visible matter. Zwicky realized that the cluster had to contain more, unseen mass to keep these galaxies bound together.

In the early 1970s, astronomers discovered that the speed of stars and clouds of hydrogen atoms rotating in a galactic disk is nearly constant all the way out to the edge of the galaxy (Fig. 1, overleaf). By Newton's law of gravitation, this implied that the amount of matter at increasing radius is not falling away — and yet starlight fades towards the edge of the galaxy. This suggested that the disk of a galaxy might be surrounded by a massive halo of dark, unseen matter. Further studies of our own Galaxy, the Milky Way, have indicated that its dark halo may extend to more than ten times the radius that is optically observed. The existence of a dark halo is also suggested by the stability of the Milky Way. When viewed edge-on, the Milky Way (and other similar spiral-arm galaxies) appears as a thin disk (shown above). This disk would be unstable against its own gravity, unless it was embedded in a more massive, sphere-like, dark halo of matter.

Modern evidence

Over the past two decades, the case in favour of dark matter has been strengthened. The evidence is becoming even more convincing as data accumulate from sky surveys, such as those performed using the Canada–France–Hawaii Telescope (CFHT) on Mauna Kea, Hawaii, and the Sloan Digital Sky Survey (SDSS) based in New Mexico. From his general theory of relativity, Einstein predicted that light from a distant

star passing close to a massive object (such as a galaxy) is deflected by the gravity of that object. The effect of such a 'gravitational lens' is generally small (unless the light passes close to the core of the lensing galaxy). But the modern surveys, taking in hundreds of thousands of galaxies, reveal statistically unambiguous lensing signals from the images of distant galaxies distorted by other galaxies closer to Earth².

When the distance to the lensing galaxy is known (as it is in the SDSS observations), the magnitude of the distortion gives a quantity known as the mass-to-light ratio (M/L) for the lensing galaxy. The mean M/L value for galaxies derived from these observations is about 150 — and yet if galaxies contained only stars, and no dark matter, the expected value would be between 5 and 10. The measured ratio also matches the value that has been inferred from optical and X-ray observations. Furthermore, when the mean M/L value is multiplied by the estimated density of light in the Universe, the resulting estimate of the total matter density ties in well with cosmological expectations.

Independent, and probably the most compelling, evidence for dark matter comes from observations of the cosmic microwave background (CMB) — the relic radiation of the Big Bang that fills the Universe. Its discovery in 1965 by Penzias and Wilson³ established modern cosmology. The CMB has a black-body temperature of 2.7 K and appears uniform across the sky. But in 1992 the Cosmic Background Explorer (COBE) satellite measured⁴ tiny fluctuations in this temperature, at the level of one part in 100,000.

The significance of this discovery is immense. It verified the model for the formation of cosmic structure — that these small fluctuations, or inhomogeneities, were the seeds for the gravitational clustering of matter into the structure of the Universe. Moreover, the magnitude of the fluctuations matches the predicted value from the structure-formation

theory that assumes the presence of dark matter: if there were no dark matter, the fluctuations should instead be about ten times as large as this.

The state of cosmology

High-resolution observations of CMB fluctuations have begun to provide detailed information for cosmological models, and a consistent theory is now emerging for the evolution of the Universe and the formation of its structure. The crucial ingredients of the theory are dark matter and the dark energy of the vacuum, which is sometimes referred to as the cosmological constant, Λ (ref. 5). We cannot construct a consistent mathematical theory without these two dark components: it seems impossible to discard the dark-matter hypothesis without endangering our whole modern scheme of cosmology.

The evolution of the Universe is governed by two equations (derived from the Einstein equation), which contain three free parameters⁶: Hubble's constant, H_0 , which represents the expansion rate of the Universe; and the matter density Ω_M and the vacuum energy density Ω_Λ , both of which are defined relative to the critical density ($0.9 \times 10^{-29} \text{ g cm}^{-3}$). The Universe was set expanding by the Big Bang. That expansion is slowed by the gravitational pull of the mass in the Universe, but is supported — even accelerated — by the vacuum energy: whether the Universe expands for ever or eventually collapses depends on the balance between mass and vacuum energy.

A powerful cosmological test is to examine whether a consistent set of parameters is derived from different observations. In recent years, progress here has been impressive. For example, the Hubble Space Telescope Key Project⁷ has pinned down the long-controversial value of H_0 more accurately than ever before, and studies of the brightness of distant supernovae suggest that the cosmological constant cannot be zero^{8,9} (Fig. 2, overleaf).

E.L. WRIGHT (UCLA)/COBE/DIRBE/NASA

The greatest advance, however, comes once again from the data on the CMB fluctuations. The CMB gives access to the state of the Universe at a time when it was only one-thousandth of its present size (defined as a 'redshift' of 1,000) — around 400,000 years after the Big Bang. This is in contrast to observations from optical astronomy, which reveal the Universe as it is today or, at best, when it was half its present size. So if the parameters derived from CMB measurements agree with those from local (optical) tests, it would be a dramatic demonstration that our understanding of structure formation and the evolution of the Universe is correct.

Several experiments^{10–13} (most recently the NASA Wilkinson Microwave Anisotropy Probe, WMAP¹⁴) have provided high-resolution maps of the cosmic temperature fluctuations. When the signal is broken down into its component frequencies, this Fourier spectrum, as it is called, shows a characteristic peak structure from which two main conclusions can be drawn. First, the total mass-energy density of the Universe is close to the critical value — that is, $\Omega_M + \Omega_\Lambda \approx 1$. The matter density (visible and dark) accounts for only 27% of the critical value, so the rest is attributed to the vacuum energy density, and thus to the cosmological constant. This is the most compelling argument for a non-zero cosmological constant, something that has for many years been anathema to physicists, especially Einstein (who introduced the constant, but later regretted it).

The second conclusion from CMB data is that the Universe has a dark-matter density, Ω_{DM} , of 0.23, and that the density of ordinary (baryonic) matter, Ω_b , is 0.04. These numbers tie in with the data from optical observations — a dramatic confirmation that the underlying models are correct. So it seems that the Universe is made up as follows: 73% dark, or vacuum, energy, 23% dark matter and 4% ordinary matter (only 0.5% is visible using optical astronomy, with another 0.5% visible at X-ray wavelengths). For some reason, most of the energy of the Universe is stored in invisible form.

WIMPs or MACHOs?

Having established the existence of dark matter in the Universe, what could it be? Might it be baryons — just ordinary matter, such as protons and neutrons? The simple atoms helium and deuterium formed around three minutes after the Big Bang¹⁵, and their measured abundance in the Universe today constrains models of nucleosynthesis, the nuclear reactions that produced the chemical elements. The nuclear-reaction rate also constrains the total abundance of baryons — and it's too low to make up the amount of dark matter needed. Furthermore, the baryon abundance derived from CMB data matches the value from calculations of primordial nucleosynthesis, and is also consistent with the

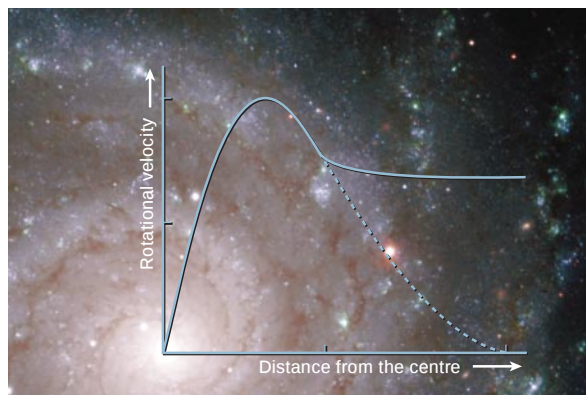


Figure 1 The flat rotation curve. From observations of starlight alone, the rotational velocity of material in the arms of a spiral galaxy (such as M74, shown here) would be expected to fall exponentially towards the outer reaches of the galaxy (dashed line). In fact the curve flattens (solid line), suggesting that the galaxy is surrounded by a halo of unseen, dark matter.

number obtained by extrapolating the local baryon distribution to the whole Universe (Fig. 3). The Universe must be dominated by some other, non-baryonic form of matter.

A non-baryonic particle postulated as a candidate for dark matter is the neutrino. Until recently there was no proof that neutrinos have mass — a vital property for dark matter, whose gravitational influence is well known. But the detection of neutrino oscillations by the SuperKamiokande¹⁶, SNO¹⁷ and KamLAND¹⁸ experiments has shown that neutrinos do indeed have mass. However, these experiments have only measured the difference in mass between the three types of neutrinos, not the absolute mass values. Making a reasonable assumption for the maximum values of these particle masses gives $\Omega_{DM} \approx 0.15$.

But in fact neutrino dark matter doesn't work. Neutrinos are classed as 'hot' dark matter, because in the early Universe they would have behaved like radiation. If the Universe were dominated by these fast-moving neutrinos, the density fluctuations in the early Universe (that eventually led to the formation of galaxies and clusters of galaxies) would have been smeared out — no model for structure formation including neutrinos as a substantial component of dark matter can reproduce the actual structure of the Universe.

What is needed is a heavy particle that behaves more like 'cold' matter than hot radiation, with a mass, say, some hundred times that of the proton — a weakly interacting massive particle, or WIMP. No such particle has been found yet, but the most promising candidate belongs to the postulated family of supersymmetric particles. Supersymmetry (SUSY) is an extra symmetry of nature, thought up by particle physicists, by which the number of elementary particles is doubled. The symmetry exists between two classes of particles, bosons (integer-spin particles) and fermions (half-integer spin particles): every boson has a fermion partner, and vice versa. So, for example, the electron (a fermion) has a bosonic partner, the super-symmetric electron, or 'selectron'; each fermionic quark is paired with a bosonic 'squark'. The bosons (such as the W, the Z and the photon) have fermionic partners called

charginos and neutralinos, this latter group being the prime candidates for dark matter. These particles are expected to have masses between 100 and 1,000 GeV (1 GeV = 10^9 electron volts, roughly the mass of a proton).

The SUSY theory was originally invented to solve a problem that arose in attempts to devise a unified theory of elementary particles and forces. But a strong argument in favour of SUSY is that it predicts a value for the ratio of the strengths of the electromagnetic and weak forces between particles (known as the weak mixing angle) that agrees precisely with experiment. The lightest supersymmetric particle might be stable against decay. If so, the predicted abundance of this particle is of the right order of magnitude to account for dark matter (using a reasonable parameter range in the particle theory).

There is also another, yet more exotic particle that is a candidate for cold dark matter. This is the 'axion'. In particle physics, interactions mediated by the weak force suffer CP violation, a distorting effect that leads to an imbalance between matter and antimatter. But interactions mediated by the strong force, and governed by quantum chromodynamics, are not subject to CP violation. The existence of the axion is postulated to explain why this is. If axions have a mass of about 10^{-5} eV, they could dominate the mass of the Universe and behave as dark matter.

But cold dark matter need not necessarily be made of elementary particles. Massive compact halo objects (MACHOs), such as low-luminosity stars or black holes, could also fit the bill. But to be viable candidates, these objects would have had to have formed before the epoch of nucleosynthesis, three minutes after the Big Bang, to evade the constraint from nucleosynthesis on baryon abundance. This condition rules out stars. The possibility of black holes as dark matter is not completely ruled out, although exotic *ad hoc* assumptions must be made about conditions in the very early Universe for them to have formed. Nevertheless, there have been interesting searches for MACHOs in galactic haloes, using the fact that they would act as gravitational lenses for light from distant stars (such as those in the Magellanic Clouds). Such gravitational-lens

effects have in fact been detected, but the frequency with which they occur is much lower than is needed for them to account for dark matter.

A further possibility is that dark matter is not actually matter at all, but a manifestation of a change in Newton's law at large distance scales: if instead of the usual $1/r^2$ dependence (where r is the distance between two gravitationally attracted objects) the gravitational force became proportional to $1/r$ at large scales, this could explain the flat rotation curve of Fig. 1 without invoking dark matter. The strength of the gravitational force itself may also increase at large scales. This hypothesis is known as modified newtonian dynamics¹⁹, or MOND. But for MOND, there is no consistent field theory of gravity, nor even a relativistic version of the theory. MOND also has no predictive power for the formation of cosmic structure, and has been all but discounted as a rival to dark matter.

Prospects

The cosmological argument in favour of dark matter is compelling, and as astronomers improve their measurements of the cosmological parameters, such as the Hubble constant and the total mass density, the existence of dark matter may be proved beyond doubt. The challenge for physicists is the direct detection of dark matter in laboratory experiments.

Several experiments are already operating, such as DAMA²⁰, CDMS²¹ and EDELWEISS²², which search for scintillation in a crystal, or ionization in germanium or silicon detectors, as signatures of dark-matter particles.

Although reluctant to interact, dark-matter particles from the galactic halo could be expected to collide with ordinary nuclei in the volume of the detector; the recoil of the nuclei from the collision could be detected through the ionization it causes. A distinctive signature would be a seasonal variation in the recoil signal, between summer and winter, caused by Earth's motion as it orbits the Sun through the dark-matter 'wind'. But so far the sensitivities of these experiments have barely reached the parameter space expected in SUSY models — much larger detectors are needed. A detector that uses nuclei with spin should also be developed: it is possible that the dark-matter particle might interact through a coupling to the spin of the detector nuclei.

Accelerator-based searches for SUSY particles are a parallel, promising effort. The Large Electron Positron collider at CERN, Geneva, failed to find any evidence for SUSY particles; nor has a likely dark-matter particle been found at the Tevatron proton-antiproton collider in operation at Fermilab in Chicago. The Large Hadron Collider, under construction at CERN and due to turn on in 2007, will continue the effort by searching for evidence of SUSY particles produced in high-energy proton-proton collisions. If stable SUSY particles are identified, their relic abundance can be calculated and translated into a cosmic mass density without much ambiguity.

The search for axions is also under way^{23,24}. In a strong magnetic field, a cosmic axion may convert into a microwave photon, which could be detected as a resonant effect in a microwave cavity. But in this area too, a

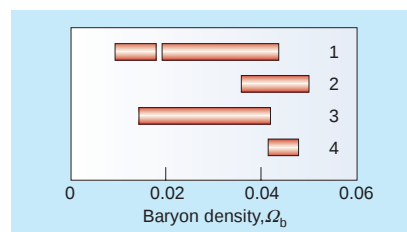


Figure 3 The cosmic baryon abundance. Measurements from different sources indicate that the value of the baryon abundance, Ω_b , is likely to fall around the 4% mark. Results 1 and 2 are derived from helium^{26,27} and deuterium²⁸ abundances, respectively, and additionally constrained by models of Big Bang nucleosynthesis. Result 3 is the value measured locally²⁹. By far the most accurately measured value (4) comes from the WMAP studies of the cosmic microwave background radiation²⁵.

considerable increase in the sensitivity of the detectors is needed for a serious search.

We have reached a basic understanding of the evolution of the Universe, and of the formation of large-scale structure within it, that relies on the presence of cold dark matter, as well as dark energy. In fairness, there are a few aspects of astrophysics that the cold-dark-matter hypothesis struggles to explain — such as small-scale structure and galaxy formation. The challenge now is to solve these problems and, once the true nature of dark matter is discovered, to complete the cosmological picture of our Universe.

Masataka Fukugita is at the Institute for Cosmic Ray Research, University of Tokyo, Kashiwa 2778582, Japan.

e-mail: fukugita@icrr.u-tokyo.ac.jp

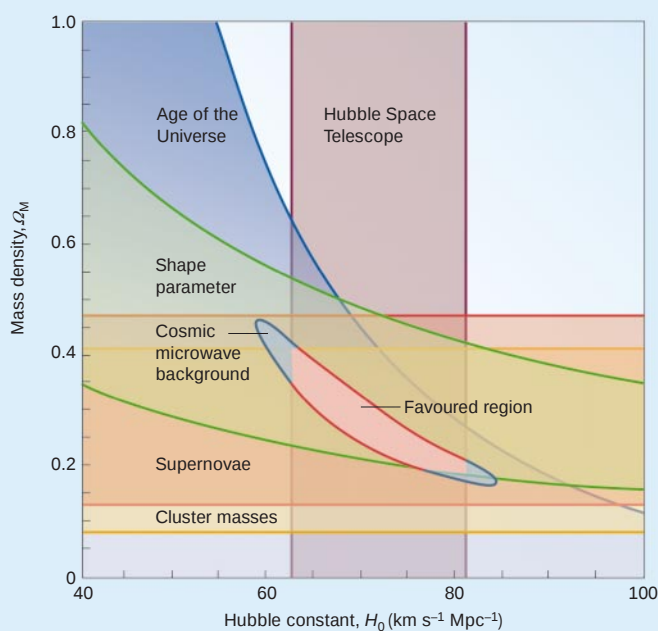


Figure 2 Consistency of the cosmological parameters. The combination of a great variety of astronomical data signals a small region (pink) in which the values of the mass density of the Universe, Ω_M , and the Hubble constant, H_0 , are expected to lie. Data come from, among others, the Hubble Space Telescope Key Project⁷, supernovae^{8,9} and most accurately of all from the cosmic microwave background²⁵. (The contours represent two-sigma statistical errors on the measurements; a flat Universe, $\Omega_M + \Omega_\Lambda = 1$, is assumed.)

1. Zwicky, F. *Helv. Phys. Acta* **6**, 110–127 (1933).
2. McKay, T. A. *et al. Astrophys. J.* **571**, L85–L88 (2002).
3. Penzias, A. A. & Wilson, R. W. *Astrophys. J.* **142**, 419–421 (1965).
4. Smoot, G. F. *et al. Astrophys. J.* **396**, L1–L5 (1992).
5. Coles, P. *Nature* **389**, 741–744 (1998).
6. Weinberg, S. *Gravitation and Cosmology* (Wiley, New York, 1972).
7. Freedman, W. L. *et al. Astrophys. J.* **553**, 47–72 (2001).
8. Riess, A. G. *et al. Astron. J.* **116**, 1009–1038 (1998).
9. Perlmutter, S. *et al. Astrophys. J.* **517**, 565–586 (1999).
10. de Bernardis, P. *et al. Nature* **404**, 955–959 (2000).
11. Hanany, S. *et al. Astrophys. J.* **545**, L5–L9 (2000).
12. Halverson, N. W. *et al. Astrophys. J.* **568**, 38–45 (2002).
13. Sivers, J. L. *et al. Preprint astro-ph/0205387* at <http://arXiv.org> (2002).
14. Bennett, C. L. *et al. Preprint astro-ph/0302207* at <http://arXiv.org> (2003).
15. Weinberg, S. *The First Three Minutes* (Deutsch, London, 1977).
16. Fukuda, Y. *et al. Phys. Rev. Lett.* **81**, 1562–1567 (1998).
17. Ahmad, Q. R. *et al. Phys. Rev. Lett.* **89**, 011301 (2002).
18. Eguchi, K. *et al. Phys. Rev. Lett.* **90**, 021802 (2003).
19. Milgrom, M. *Sci. Am.* **287**, 42–52 (August 2002).
20. Bernabei, R. *et al. Phys. Lett. B* **480**, 23–31 (2000).
21. Abusaidi, R. *et al. Phys. Rev. Lett.* **84**, 5699–5703 (2000).
22. Benoit, A. *et al. Phys. Lett. B* **513**, 15–22 (2001).
23. Hagmann, C. *et al. Phys. Rev. Lett.* **80**, 2043–2046 (1998).
24. Yamamoto, K. *et al. Preprint hep-ph/0101200* at <http://arXiv.org> (2001).
25. Spergel, D. N. *et al. Preprint astro-ph/0302209* at <http://arXiv.org> (2003).
26. Olive, K. A., Steigman, G. & Walker, T. P. *Phys. Rep.* **389**, 333–334 (2000).
27. Izotov, Y. I. & Thuan, T. X. *Astrophys. J.* **500**, 188–216 (1998).
28. Kirkman, D. *et al. Preprint astro-ph/0302006* at <http://arXiv.org> (2003).
29. Fukugita, M., Hogan, C. J. & Peebles, P. J. E. *Astrophys. J.* **503**, 518–530 (1998).

Are fast-moving elephants really running?

Despite their unseemly bulk, elephants can hit high speeds — but use an unusual style.

It is generally thought that elephants do not run^{1–5}, but there is confusion about how fast they can move across open terrain and what gait they use at top speed. Here we use video analysis to show that Asian elephants (*Elephas maximus* L.) can move at surprisingly high speeds of up to 6.8 m s^{-1} (25 km h^{-1}) and that, although their gait might seem to be a walk even at this speed, some features of their locomotion conform to definitions of running.

Elephants moving rapidly have been estimated^{2–4} to reach speeds of about 4 m s^{-1}

(15 km h^{-1}), although anecdotal evidence¹ claims that they can reach 11 m s^{-1} (40 km h^{-1}). To investigate the gait used by elephants at top speeds, we used video analysis to study 42 healthy, active Asian elephants throughout Thailand (for details, see supplementary information). The skin was marked with non-toxic tempera paint dots over the limb-joint centres of the elephants (right fore- and hindlimbs; Fig. 1), estimated by palpation and manipulation of the joints; these dots were used for later video digitizing. Mahouts guided the elephants along a 30-metre course, parallel to the field of view of the video camera (60 Hz). Elephants had at least 10 m to accelerate or decelerate before and after this 30-m course. A total of 188 trials were carried out; trials with sudden accelerations or decelerations in the 10-m videotaped stretch of the course were omitted.

Our digitization of the hip-joint markers (using Peak Motus, Peak Performance, Colorado) measured the average velocity along the central 10 m of the track. We used the length of the thigh segment (between the centres of hip- and knee-joint markers) to scale the digitized video coordinates to real dimensions. Photocell timers at either end of the track gave an average velocity across the entire 30 m, which was used as a preliminary gauge of which elephants were the fastest, as well as for comparison with the 10-m velocity to monitor speed changes. Speeds over the 10-m and 30-m courses were generally similar, indicating that the elephants did not suddenly speed up or slow down.

Of the elephants, 32 reached top speeds of over 4.0 m s^{-1} , 20 exceeded 5.0 m s^{-1} , and three attained speeds greater than 6.0 m s^{-1} . The fastest gait used by elephants has been variously described as a walk, amble, trot, pace, rack or a running walk^{1–5}, but — given that these speeds are relatively fast — how well does this gait of the fastest elephants fit the definitions of running?

Several kinematic factors distinguish quadrupedal walking from running. First, trotting and galloping are running gaits with footfall patterns that are distinct from walking⁵. Second, an aerial phase (a period during which no foot touches the ground) often marks the transition from walking to running⁵. Third, a run has been defined as any gait with a duty factor (the fraction of a complete sequence of footfalls for which a given foot is in contact with the ground) of less than 0.50 (ref. 5). Our elephants maintained the same walking footfall pattern (Fig. 2a) and always kept at least one foot in contact with the ground, although they used duty factors as low as 0.37.

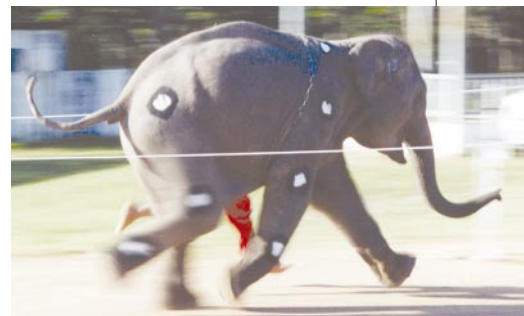


Figure 1 An Asian elephant marked with dots for gait analysis.

R. LAIR

Walking and running can also be distinguished by the forces involved. The Froude number (Fr) is a dimensionless speed⁶ calculated as $\text{velocity}^2 / (\text{acceleration of gravity} \times \text{hip height})$. At speeds beyond $\text{Fr} 1.0$, the force needed to keep the body mass on a circular arc during stance exceeds the force of gravity. Theoretically, this requires a walking animal to leave the ground and run. Most animals usually switch from a walk to a run at $\text{Fr} \approx 0.50$, presumably because of an energetic or mechanical trigger⁷, and quadrupeds switch from a trot to a gallop at $\text{Fr} \approx 2.5$ (ref. 6). The elephants routinely exceeded $\text{Fr} 1.0$, reaching Fr values as high as 3.4 — speeds that are inconsistent with a quadrupedal walking gait. Other animals, such as running birds⁸, also have non-aerial gaits at high Froude numbers.

It is the exchange of gravitational potential and kinetic energy that fundamentally distinguishes walking and running^{9–11}. The centre of mass is highest at mid-stance in walking, but lowest at mid-stance in running¹². Our analysis shows that at low speeds, as expected for walking, elephants' shoulder and hip joints rise and then fall (indicating vertical motion of the centre of mass) during the stance phase. At the highest speeds, the vertical movements of the shoulder indicate walking, but hip motion indicates running (Fig. 2b). During the stance phase of the forelimb, shoulder motion resembles walking, moving upwards and then downwards while the front foot is on the ground, whereas the hip's motion during the stance phase of the hindlimb is characteristic of running, moving downwards and then upwards.

Usually, the various criteria for walking and running are consistent, making it relatively easy to distinguish walking from running, but this is not true in the case of elephants. Our observations suggest that, at greater speeds, elephants do more than merely walk. Ground-reaction-force data are needed to demonstrate conclusively whether the elephants' gait involves spring-

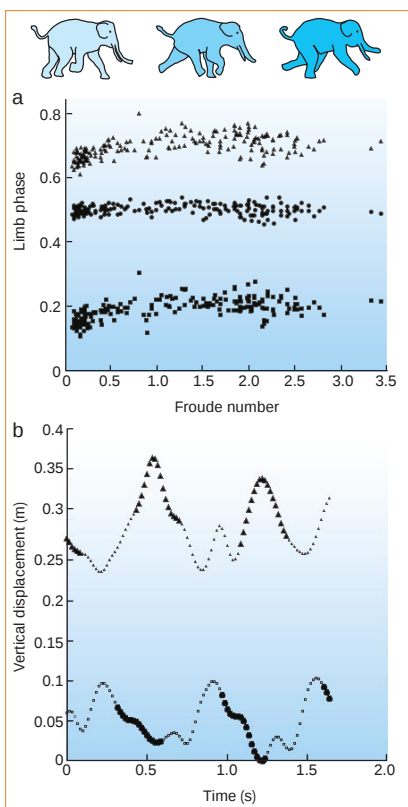


Figure 2 Biomechanics of rapid movement by Asian elephants (*Elephas maximus* L.). **a**, Relative limb phase (fraction of stride time for which each foot comes into contact with the ground after the left hind foot) plotted against dimensionless speed (Froude number). At all speeds, the left hind foot, then the left front (squares), then the right hind (circles), then the right front foot (triangles) contact the ground in the same sequence. Images of elephants moving at 6.6 m s^{-1} (Froude number, 3.1; see movie in supplementary information) illustrate these footfall patterns, which are typical of quadrupedal walking⁵. **b**, Vertical displacements of the hip (circles) and shoulder (triangles) joints plotted against time for one individual moving at 6.8 m s^{-1} (Froude number, 2.8). Large symbols indicate stance phase; small symbols indicate swing phase. This downward hip movement was typical of all fast-moving elephants; many showed a downward-then-upward pattern during stance. Roughly 2.5 strides are shown, beginning just before and ending just after the 10-m section of the observation course.

like running kinetics^{9–12} and to understand why elephants avoid using an aerial phase.

John R. Hutchinson*, Dan Famin†, Richard Lair‡, Rodger Kram§

*Biomechanical Engineering Division, Stanford University, Stanford, California 94305-4038, USA
e-mail: jrutch@stanford.edu

†School of Veterinary Medicine, University of California, Davis, California 95616, USA

‡Thai Elephant Conservation Center, Lampang 52000, Thailand

§Department of Kinesiology and Applied Physiology, University of Colorado, Boulder, Colorado 80309-0354, USA

1. Howell, A. B. *Speed in Animals* (Chicago Univ. Press, 1944).
2. Muybridge, E. *Animals in Motion* (Dover, New York, 1957).
3. Gambaryan, P. P. *How Mammals Run* (Wiley, New York, 1974).
4. Alexander, R. McN. et al. *J. Zool. Lond.* **189**, 135–144 (1979).
5. Hildebrand, M. *Am. Zool.* **20**, 255–267 (1980).
6. Alexander, R. McN. & Jayes, A. S. *J. Zool. Lond.* **201**, 135–152 (1983).
7. Farley, C. T. & Taylor, C. R. *Science* **253**, 306–308 (1991).
8. Gatesy, S. M. *J. Morphol.* **240**, 115–125 (1999).
9. Dickinson, M. H. et al. *Science* **288**, 100–106 (2000).
10. Cavagna, G. A. et al. in *Scale Effects in Animal Locomotion* (ed. Pedley, T. J.) 111–125 (Academic, New York, 1977).
11. Heglund, N. C., Cavagna, G. A. & Taylor, C. R. *J. Exp. Biol.* **97**, 41–56 (1982).
12. McMahon, T. A. et al. *J. Appl. Physiol.* **62**, 2326–2337 (1987).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Cell biology

Developmental predisposition to cancer

Many human cancers occur in renewing epithelial tissues, in which cellular lineages typically go through two distinct phases: early in life, cell populations expand exponentially to form the tissue, and for the remainder of life, the tissue is renewed by stem cells dividing to create an almost linear cellular history¹. Here we use a simple mathematical model to show that mutations that arise during the exponential phase probably seed tissues with stem cells carrying mutations that may predispose to cancer. Susceptibility to late-life cancers, such as those of the skin and colon, may therefore be influenced by somatic mutations that occur during early development.

Mutations accumulate during exponential cellular growth according to the Luria–Delbrück distribution². Let u_e be the mutation probability per cell division during exponential growth, and N the number of stem cells produced during development to seed the tissue. Starting with one cell, the number of cell divisions required to produce N cells is $N - 1$, and the total number of mutation events during the cellular history is $M = u_e(N - 1) \approx u_e N$. (The phases of cellular growth are shown in Fig. 1a.)

Figure 1b shows the probability distribution for the number of stem cells that carry mutations. These are the mutations that arise during exponential growth, before further division of the stem lineages to renew the local tissue. The average frequency of stem cells with mutations is $\bar{x} \approx u_e \ln(N)$ (ref. 2).

Although the frequency of stem cells that initially contain a mutation may be small, those mutations can contribute substantially to the total risk, R_T , of cancer. Suppose that k rate-limiting mutations are needed for cancer to develop^{3,4}, then the total risk is $R_T = N(1 - x)R_k + NxR_{k-1}$, where x is the frequency of stem cells that start with one mutation, and R_k is the risk that a particular stem-cell lineage acquires k mutations during

the phase of linear division and tissue renewal. The risk, R_k , is given roughly by the gamma distribution, which describes the probability of the occurrence of the k th event over a particular time interval. From the gamma distribution, $R_k \approx (u_s \tau)^k / k!$, where u_s is the mutation rate per stem-cell division, and τ is the total number of stem-cell divisions.

The expected fractional increase in cancer risk arising from stem cells that begin with one mutation can be calculated as $F = \bar{x} R_{k-1} / R_k \approx u_e \ln(N) k / u_s \tau$. The next step is to assign approximate magnitudes to these quantities. We can take $k \approx 5$ for the number of rate-limiting mutations required to cause epithelial cancer in humans^{3,4}, $\ln(N) \approx 20$, and τ to range from 100–1,000. This gives the fractional increase in cancer risk, F , ranging from $u_e / 10u_s$ to u_e / u_s .

If mutations accumulate with the same probability per cell division during exponential growth and linear stem-cell division ($u_e = u_s$), then the increase in risk from mutations arising in development ranges from 10–100%. If, as has been claimed⁵, mutation rates in stem cells are much lower than those during exponential growth, ($u_s \ll u_e$), then almost all cancer arises from predisposed stem-cell lineages that were mutated during development.

The risk of cancer from somatic mutations during development could be quantified by studying inbred rodents. The number of predisposing mutations per individual will vary according to the Luria–Delbrück distribution (Fig. 1b). Individuals with many predisposing mutations are likely to develop multiple independent tumours relatively early in life, whereas those with few predisposing mutations should develop few tumours relatively late in life. Controlled experiments have shown variability in cancer susceptibility among inbred rodents⁶, but the causes of this variability have not been determined.

Mutations during development seed a young individual with a small fraction of mutated stem cells. Those relatively few mutated lineages may therefore be responsible for a substantial proportion of the

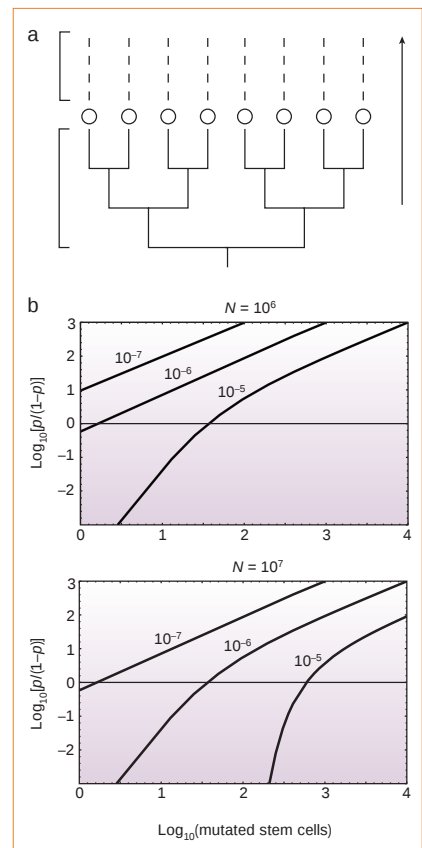


Figure 1 The role of tissue architecture in the accumulation of mutations during development. **a**, The phases of cellular growth in epithelial tissues. Cell populations increase exponentially during development, shown by a branching phase of division; at the end of development, stem cells differentiate in each tissue compartment (circles); stem cells renew each compartment by dividing to form a nearly linear cellular history (dashed lines) — each stem-cell division gives rise to one daughter stem cell that continues to renew the tissue and to one daughter transit cell that divides rapidly to produce a short-lived lineage that fills the tissue. **b**, The cumulative probability, p , for the number of stem cells mutated during development. By plotting $\log_{10}[\mu/(1-p)]$, the zero line gives the median of the distribution. N , number of stem cells produced during exponential growth. The number above each line is u_e , the mutation probability per cell added to the population during exponential growth. For a single gene, u_e is probably of the order of 10^{-7} ; thus, if there are 100 genes for which initial mutations can influence the progression to cancer, then u_e is roughly of the order of 10^{-5} . The Luria–Delbrück distribution plotted here was calculated using equation (8) of ref. 7.

cancers that develop later in life.

Steven A. Frank*, Martin A. Nowak†

*Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525, USA

e-mail: safrank@uci.edu

†Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA

1. Cairns, J. *Nature* **255**, 197–200 (1975).
2. Zheng, Q. *Math. Biosci.* **162**, 1–32 (1999).
3. Knudson, A. G. *Proc. Natl Acad. Sci. USA* **90**, 10914–10921 (1993).
4. Hanahan, D. & Weinberg, R. A. *Cell* **100**, 57–70 (2000).
5. Cairns, J. *Proc. Natl Acad. Sci. USA* **99**, 10567–10570 (2002).
6. Prehn, R. T. *Science* **190**, 1095–1096 (1975).
7. Zheng, Q. *Math. Biosci.* **176**, 237–252 (2002).

Competing financial interests: declared none.

Egg recognition and counting reduce costs of avian conspecific brood parasitism

Bruce E. Lyon

Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, California 95064, USA

Birds parasitized by interspecific brood parasites often adopt defences based on egg recognition but such behaviours are puzzlingly rare in species parasitized by members of the same species. Here I show that conspecific egg recognition is frequent, accurate and used in three defences that reduce the high costs of conspecific brood parasitism in American coots. Hosts recognized and rejected many parasitic eggs, reducing the fitness costs of parasitism by half. Recognition without rejection also occurred and some hosts banished parasitic eggs to inferior outer incubation positions. Clutch size comparisons revealed that females combine egg recognition and counting to make clutch size decisions—by counting their own eggs, while ignoring distinctive parasitic eggs, females avoid a maladaptive clutch size reduction. This is clear evidence that female birds use visual rather than tactile cues to regulate their clutch sizes, and provides a rare example of the ecological and evolutionary context of counting in animals.

Avian brood parasitism provides a model system for studying evolutionary aspects of animal cognition because recognition mechanisms are central to host defences against parasitism, and the evolutionary costs and benefits of these mechanisms can be quantified^{1–7}. Egg recognition and rejection is a particularly common defence in hosts parasitized by interspecific avian brood parasites^{2,6,7}. Recent evidence reveals that brood parasitism within species is also widespread in birds^{7–11}, but host defences based on conspecific egg recognition are surprisingly rare^{7,11}. The reasons for this absence are unclear, but hypotheses include lower costs of parasitism for hosts of conspecific brood parasites and hence reduced natural selection for defences¹¹, or insufficient variation in egg features among conspecific females for accurate egg recognition to be possible¹⁰.

Here I describe the results of a study of conspecific brood parasitism in the American coot (*Fulica americana*), an aquatic rail, which revealed high levels of accurate egg recognition. By examining how coots have incorporated egg recognition into a variety of defensive tactics to mitigate the costs of brood parasitism by conspecifics, I show that a full understanding of brood parasitism and host defences contributes to, and indeed integrates, three disparate disciplines—cognition, physiology and life-history evolution.

Parasitism is costly to hosts

Conspecific brood parasitism was both frequent, and costly to hosts, in the study population of coots near Riske Creek, British Columbia, Canada^{12,13}. During the four-year study (1987–90) 13% of all eggs in the population were laid parasitically, 41% of 417 pairs were parasitized, and hosts received an average of 3.1 parasitic eggs¹², a substantial number relative to the average clutch size (excluding parasitic eggs; 8.1 ± 0.09 eggs, $n = 388$; all values reported with means are standard errors). Indirect evidence indicates that successful parasitism was very costly to hosts owing to intense competition between host and parasitic chicks for limited parentally supplied food early in life. Chick starvation was ubiquitous during the study: 98% of all nests ($n = 177$) lost at least one chick, and on average, each nest lost 52% of its chicks, primarily to starvation¹². Although parasitism increased the total number of chicks that hatched at host

nests, it had no effect on the total number fledged, owing to complete compensatory chick mortality¹⁴. Together, these patterns indicate that each successful parasitic chick survives at the expense of a host chick—a one-for-one substitution¹⁴. This high cost per successful chick, combined with the high frequency of parasitism, should result in strong natural selection for effective host defences to reduce the costs of parasitism.

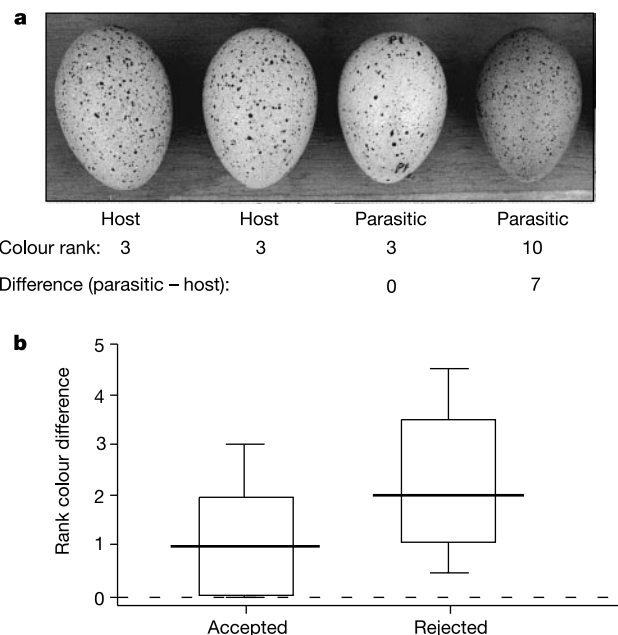


Figure 1 Egg rejection on the basis of colour. **a**, Host and parasitic eggs from a single nest, illustrating variation in rank background colour (see Methods) and the difference in ranked colour between parasitic and host eggs (all host eggs in this clutch were rank 3). **b**, Rejected parasitic eggs differed more from their host eggs in rank background colour (median difference = 2.0, $n = 28$ host–parasite dyads, see Methods) than did accepted parasitic eggs (median difference = 1.0, $n = 55$ dyads; Mann–Whitney U -test, one-tailed $P = 0.003$). Thick line on box plots, the median; box, the 25th and 75th percentiles; thin line, 10th and 90th percentiles.

Egg recognition reduces costs of parasitic chicks

Hosts reduced potential competition from parasitic chicks by rejecting parasitic eggs from their nests before they hatched (Table 1). Overall, 57 of 133 hosts (42.9%) rejected at least one parasitic egg. Egg rejection was highly non-random, indicating a specific defence against parasitism based on egg recognition (matched-pair comparison of the proportion of host and parasitic eggs rejected at each host nest; parasitic eggs rejected at a higher rate than host eggs at 50 of 56 nests, ties excluded, sign test $P < 0.0001$). Egg features such as colour and spotting patterns vary greatly among females^{12,13} and could thus serve as cues for accurate egg recognition. Rejected parasitic eggs differed more from their host eggs in background colour than did accepted parasitic eggs (Fig. 1), strongly suggesting that hosts use visual cues to distinguish parasitic eggs from their own eggs.

For hosts, egg rejection yields a high fitness benefit yet incurs only a low cost. Hosts rejected 45% of the 208 parasitic eggs laid early enough in the host's cycle to pose a threat (that is, laid within a day of clutch completion), thereby reducing the costs of parasitism by almost half. The cost of rejection to hosts—the mistaken rejection of their own eggs ("recognition error"³)—is typically measured as the frequency of all rejected non-parasitic eggs. This may not be appropriate, however, because coots have two distinct methods of egg rejection (Table 1), only one of which appears to be directed at parasitism. Most parasitic eggs were rejected by burial, while most non-parasitic eggs were rejected by outright ejection from the nest (Table 1), which suggests that the two rejection methods solve different problems.

All eggs that were known to be cracked or rotten were ejected rather than buried (7 non-parasitic, 4 parasitic), whereas none of the many buried eggs recovered from nesting material were ever damaged or rotten. This observation suggests that ejection is used to quickly remove damaged or rotten eggs that threaten the entire clutch and is not a defence against parasitism. Moreover, ejection of non-parasitic eggs was not disproportionately common at parasitized nests, as would be expected were it a defence against parasitism (ejection at 19 parasitized and 24 unparasitized nests, χ^2 goodness-of-fit test based on a 41% parasitism rate, $\chi^2 = 0.18$, $P = 0.67$). Thus, the cost of rejection to hosts is probably not the frequency of all rejected non-parasitic eggs, but the frequency with which non-parasitic eggs are buried (Table 1, 9/3,062 eggs = 0.29%), a very low cost. The idea that burial and ejection solve different problems, using different cues, is supported by the observation that many species that do not recognize or reject parasitic eggs readily eject damaged eggs (whether host or parasitic)¹⁵. Thus, unlike burial, ejection does not require recognition of foreign eggs.

Hosts exhibited a second, subtler defence that involved recognition without outright rejection—parasitic eggs were banished to the periphery of the clutch where incubation positions are likely to be inferior. A matched-pair comparison of the proportion of all parasitic and host egg positions that were in outer and inner positions, respectively, at each nest revealed that host eggs were in central positions proportionately more often than parasitic eggs at 22 of 25 nests (sign test $P = 0.0002$; median proportion in central

positions 13.3% for parasitic eggs, 22.6% for host eggs). Because of these positional effects, parasitic eggs took longer to hatch than the corresponding host eggs laid on the same day (Fig. 2). Hosts benefit from this outcome because chick survival is strongly linked to relative hatching order¹², so delaying the hatching of parasitic chicks reduces their survival and thus impact on host chicks.

These patterns indicate some level of recognition, so I wondered why the eggs were not rejected outright. One possibility is that hosts use incubation positions to deal with eggs that are distinguished as parasitic eggs with less certainty, and hence entail a higher risk of a recognition error. While positional effects are less effective than outright rejection, the cost of making a mistake and banishing a host egg to the outside is also less extreme than the outright rejection of a host egg. Regardless, the addition of egg recognition to a behaviour that is widespread in birds, shuffling eggs to ensure equitable incubation conditions, has resulted in an unusual defence against brood parasitism. Moreover, the observation that not all recognition leads to rejection means that using egg rejection as a proxy for egg recognition is not a reliable measure⁴.

Why coots?

Conspecific egg recognition and rejection are known to be well developed in only three taxa: in coots and related rails^{13,16,17}, weaverbirds (genus *Ploceus*)¹⁸ and the ostrich (*Struthio camelus*)¹⁹. The evolution of defences against parasitism will depend not only on the cost of each successful parasitism but also on the frequency of such parasitism—that is, population level fitness costs. Estimates indicate that in the absence of host defences, parasitism would cause a 5.4% reduction in the total population production of non-parasitic coot chicks¹¹ (based on estimating the number of parasitic chicks that would survive in the absence of host defences (see Methods) and subtracting one host chick from the population for each surviving parasitic chick). Similar cost estimates are not available for any other conspecific parasite, but comparison with hosts parasitized by a well-known interspecific parasite, the common cuckoo (*Cuculus canorus*), is informative.

Hosts successfully parasitized by cuckoos raise no chicks of their own⁷, so the frequency of parasitized nests provides a maximum estimate for the population level costs that would be incurred without egg rejection (assuming that all cuckoo eggs are successful). Davies and Brooke²⁰ report parasitism frequencies for each of 14 host species in the United Kingdom, and the above fitness cost for coots equals the cost of parasitism for the most heavily parasitized cuckoo host, the reed warbler (*Acrocephalus scirpaceus*): both would lose about 5% of fitness to parasitism without defences. This is an unexpected finding, given the enormous cost to hosts of raising a cuckoo chick rather than a coot chick, but it can be attributed to the higher frequency of parasitism in coots than in cuckoo hosts. The extreme fitness costs for coots helps explain the evolution of their sophisticated battery of defences based on egg recognition.

Egg recognition and life-history evolution

Hosts also used recognition to deal with a previously unappreciated cost of parasitism, a maladaptive clutch-size reduction caused by misinformation provided by parasitic eggs. Coots, like many birds,

Table 1 Frequency and method of rejection of parasitic and non-parasitic eggs

Type of egg	Number of eggs		Percentage of eggs rejected	Number of eggs rejected by*		Percentage of rejections by burial
	Rejected	Not rejected		Burial	Ejection	
Parasitic	174	356	32.8%	142	32	81.6%
Non-parasitic	73	3,062	2.4%	9	65	12.2%

Rejected non-parasitic eggs include 38 eggs in 24 parasitized nests and 35 eggs in 25 unparasitized nests.

*When nests are counted only once for each egg type (parasitic or not) or rejection method (burial or ejection), burial was the method used for 75.9% of 83 nests rejecting parasitic eggs, and for 12.5% of 48 nests rejecting non-parasitic eggs; $\chi^2 = 49.0$, $P < 0.0001$.

are 'indeterminate' egg-layers that use an external cue, such as the number or surface area of eggs in the nest, to cease further development of egg follicles and thereby regulate their clutch sizes²¹. With this mechanism, the addition or removal of eggs early in the laying cycle alters the number of eggs present in the nest (the termination cue) relative to the number actually laid by the female up to that point, causing the female to alter the total number of eggs she lays. Thus, parasitic eggs laid early in the host's laying cycle could affect clutch-size decisions by decreasing the total number of eggs laid by the host.

Using a titration approach to identify the temporal window of sensitivity and the degree of response to the extra eggs added by parasites (see Methods), I found that hosts do alter their clutch size in response to parasitic eggs, and that the strongest effect was observed for eggs added by the host's third day of laying. I then separately examined the clutch-size responses of two classes of host nests—nests where the parasitic eggs were accepted and nests where they were rejected. The acceptors reduced their clutch size in response to early-laid parasitic eggs (Fig. 3a: slope = -1.06 ± 0.24 host eggs per parasitic egg added; $F_{1,111} = 20.18$, $P < 0.0001$). In contrast, the clutch sizes of hosts that rejected parasitic eggs were unaffected by the number of early-laid parasitic eggs (Fig. 3b: slope = -0.03 ± 0.29 ; $F_{1,111} = 0.013$, $P = 0.91$). Ignoring the number of parasitic eggs received and comparing means revealed that clutch sizes differed significantly between these two classes of nests (6.65 ± 0.45 eggs for 17 acceptors, 8.12 ± 0.42 eggs for 17 rejecters; $t = 2.41$, $P = 0.02$).

Why did early-parasitized acceptor and rejecter females differ in their clutch-size responses? The difference is not explained by inherent differences in female quality between acceptor and rejecter nests because there was no difference in the clutch sizes of rejecter and acceptors parasitized after host clutch-size decisions were made (that is, reference hosts in Fig. 3: 8.18 ± 2.46 eggs for 39 acceptors, 8.12 ± 0.32 eggs for 17 rejecters; $t = 0.14$, $P = 0.89$). The difference is also not explained by the effects of rejection itself on the total number of eggs present when hosts made their clutch-size decisions because rejecters did not reject the parasitic eggs until after they had made their clutch-size decisions (Fig. 3c). Consequently acceptor and rejecter nests had the same number of total eggs (host plus parasite) on day 3, the end of the responsive period for clutch-size adjustment (4.67 ± 0.21 eggs for acceptors, 4.50 ± 0.20 eggs for rejecters; $t = 0.57$, $P = 0.57$), an observation that rules out the use of tactile cues to regulate clutch size²¹.

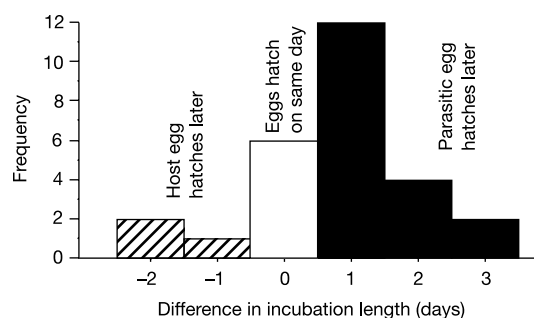


Figure 2 Difference in the length of the incubation period for paired parasitic and host eggs laid in the same nest on the same day. The incubation period is the time from laying to hatching. Black bars indicate pairs of eggs where the parasitic egg took longer to hatch, striped bars where the host egg took longer, and white bars where both eggs hatched on the same day. On average, parasitic eggs hatched 0.78 days later than their corresponding host egg (matched-pair comparison of average values for each host–parasite female dyad, rather than individual pairs of eggs; $t = 2.63$, degrees of freedom = 16, one-tailed $P = 0.009$).

Coots can count

The different clutch-size responses of acceptors and rejecters appears to arise instead from their use of visual cues—specifically counting the number of eggs perceived as their own—to decide when to stop further development of maturing egg follicles. The perception of such visual information would have differed for acceptors and rejecters, as follows. By chance, acceptor females received parasitic eggs too similar to their own to permit recognition (or rejection, Fig. 1) so they included the parasitic eggs in their count during clutch-size assessment. The observed clutch-size reduction, 1.06 fewer host eggs per parasitic egg (Fig. 3a), does not differ significantly from a slope of -1.0 ($P > 0.05$), the slope expected if a host counts each parasitic egg as one of her own. At rejecter nests, in contrast, parasitic eggs were sufficiently different to be recognized (Fig. 1), and their lack of impact on clutch size (Fig. 3b) suggests that rejecter females counted only their own eggs and ignored parasitic eggs when making their clutch-size decisions.

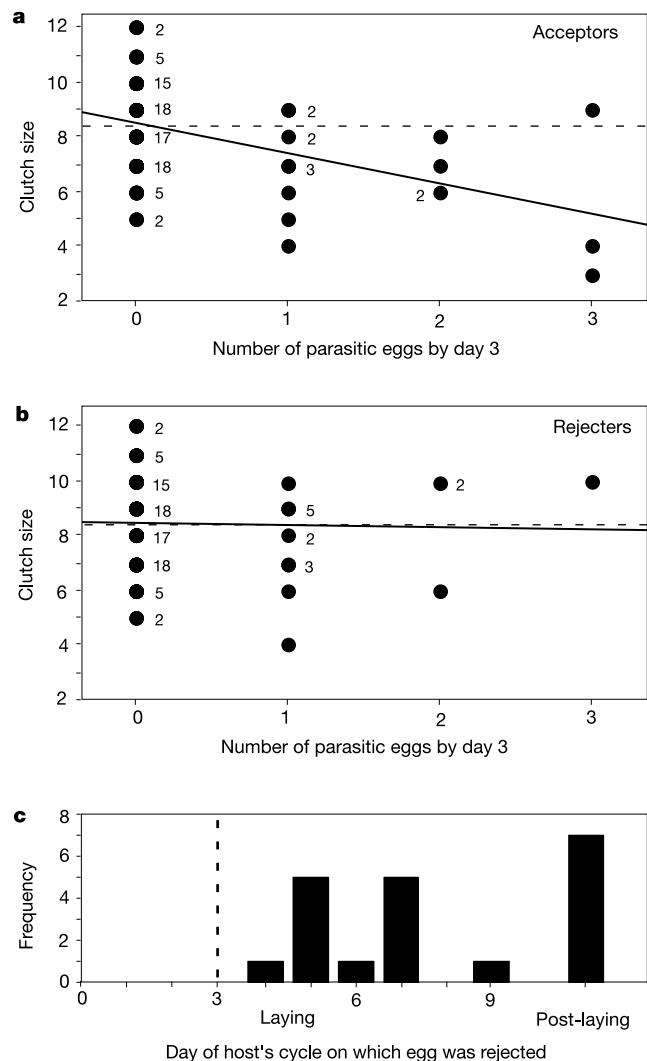


Figure 3 Host clutch-size response to early-laid parasitic eggs. **a, b**, Parasitic eggs includes those added during the host's responsive period (up to day 3 of laying cycle) and either accepted (**a**) or rejected (**b**) by the hosts. Solid lines, simple regression lines; dashed lines, mean clutch size for reference birds not parasitized early. Numbers indicate data points with multiple identical values. **c**, The timing of egg rejection for the nests used to examine the clutch-size responses of rejecters (shown in **b**). All eggs were rejected after the three-day window of host responsiveness (vertical dashed line) during which clutch-size decisions are made.

The ability of females to count only their own eggs in a mixture of eggs is a remarkable feat that provides a convincing, rare example of counting in a wild animal¹. This observation also has broad implications for proximate mechanisms of clutch-size regulation, and their evolutionary consequences. These clutch-size patterns provide the first convincing evidence that birds use visual cues and egg recognition to regulate their clutches, because tactile cues cannot explain the different responses of acceptors and rejecters. The widespread assumption of the ubiquity of tactile information as decision cues for indeterminate egg layers²¹ clearly needs to be reassessed.

The different clutch-size responses to parasitic eggs at acceptor and rejecter nests (Fig. 3) also raise questions about the adaptive basis of these patterns. Theory predicts that hosts can benefit from clutch-size reductions in some situations^{22,23}. However, none of these benefits apply to the females who rejected eggs because egg rejection nullifies any costs that would favour a smaller clutch size. For these females, avoiding an inadvertent clutch-size reduction, as they did (Fig. 3b), is the adaptive response, a response that required the evolution of a sophisticated mechanism consisting of recognizing and counting the right eggs, while discounting parasitic eggs. Further work is needed to determine if the clutch-size reductions at coot nests where eggs are accepted (Fig. 3a), and in other species in general²², is a selected, adaptive response favoured by natural selection, or an inadvertent, maladaptive artefact of the mechanism used to regulate clutch size.

Experiments have confirmed the existence of simple counting abilities in a diversity of taxa in the laboratory^{1,24}. However, the ecological and evolutionary context of such capacities is unknown, and clear examples of counting in wild animals are virtually non-existent^{1,25}. In American coots, egg counting is directly linked to clutch size, a key life-history trait with fitness consequences. Visual egg counting may turn out to be common in a variety of birds that use external cues to regulate their clutch size, but without the clear signal provided by brood parasitism and egg recognition it may prove difficult to detect in observational field studies. More broadly, a connection between counting abilities and reproductive investment is likely to be widespread among animals with parental care²⁶, because fitness is tightly coupled with the number of offspring produced, and parents should benefit from mechanisms that enable them to fine-tune reproductive decisions to fitness payoffs²⁶. Studies of reproductive decisions concerning numbers of eggs and offspring can provide fertile ground for studying counting in animals and may provide model systems for integrating animal cognition and physiological mechanisms in an ecological and evolutionary context.

Methods

Detecting parasitism and parasitic eggs

Depending on the year of the study, each nest was checked daily ($n = 206$) or every second day ($n = 211$). All new eggs were individually numbered with indelible felt pens. Parasitism was detected primarily by the appearance of more than one new egg in a 24-h period but egg features were then used to determine which of the new eggs were parasitic, and to match parasitic eggs laid by the same female (parasites often laid several eggs per host nest and many host nests were parasitized by several females¹³). The accuracy of these field techniques has been verified by both discriminant function analysis based on egg features¹³ and DNA fingerprinting¹⁴.

Egg features, incubation positions and rejection

On each nest visit all eggs were censused to determine whether new eggs had been laid or previously labelled eggs had disappeared. Eggs that disappeared were assigned to two rejection categories: (1) buried, if recovered from the nest material or last observed at least 50% buried (most buried eggs were initially observed partly buried); and (2) ejected, for all other eggs that disappeared without meeting the criteria for burial. Ejection is thus a catch-all category that may include sources of egg loss unrelated to rejection, including partial predation and accidental displacement during parasitism (parasites seem not to remove host eggs deliberately¹³). In several cases, ejected eggs were recovered from the water below nests.

To determine whether coots non-randomly keep parasitic eggs to the periphery of the clutch during incubation, on each nest visit the position of each host and parasitic egg was scored as either 'central' (egg completely surrounded by other eggs) or 'outer' (egg

lacked neighbouring eggs on at least one side). Coots shuffle egg positions at least daily so eggs were not constrained to particular positions. For each nest I then pooled positions from all visits and determined the proportion of host and parasite egg positions that were central and outer, and then did a matched-pair comparison for all nests comparing the proportion of parasitic and host egg positions at each nest that were in outer positions.

To determine whether egg features affected egg rejection, I scored the background colour of eggs on a ranked darkness scale that ranged from 1 (white) to 10 (dark brown; see Fig. 1) by comparing them in the field to a reference collection of ten real coot eggs. Most nests in the analysis are represented by a single host-parasite dyad (61 nests), but for the nine hosts parasitized by several parasites, contrasts were included for each of the host-parasite dyads at the nests. Five of the nine nests parasitized by several females rejected eggs from some of the parasites but not others, indicating that rejection among different dyads involving the same host is independent. For each host-parasite contrast, I computed the difference between the mean background egg rank of the parasitic eggs and each of the host eggs, and then used the smallest value as the index of difference between the host and parasite. A host-parasite dyad was scored as 'reject' if at least one parasitic egg was rejected, but in fact rejection was all or none for most dyads. I also omitted acceptors that received parasitic eggs late in incubation (day 7 or later) because there may have been insufficient time for the birds to reject eggs. In all statistical comparisons, data points were nests or females, not individual eggs.

Fitness costs of parasitism

To estimate the number of parasitic chicks that would survive in the absence of egg rejection, each rejected parasitic egg laid in a successful host nest was assigned a survival probability based on its predicted position in the hatching order (based on laying order) using an empirically derived relationship for survival probability relative to hatching order¹².

Clutch-size comparisons

To determine the temporal window over which hosts adjusted their clutch sizes in response to parasitism, I conducted a series of linear regressions, each comparing host clutch size (dependent variable) with the number of parasitic eggs received by a specific cut-off day in the host's laying cycle (independent variable; assessed day 3, 4 and 5 of host's cycle, respectively). To avoid spurious effects due to female quality, only parasitized birds were used in these analyses, and 82 birds not parasitized early (that is, parasitized on day 6 or later) served as the reference group (that is, 0 parasitic eggs by day 3) for determining a clutch-size response of early-parasitized birds. Clutch size showed the strongest decline with number of parasitic eggs received by the host's third day of laying (slope for day 3 regression = $-0.63 (\pm 0.20)$, $P = 0.002$, $n = 21$ nests parasitized by day 3; slope for day 4 regression = $-0.39 (\pm 0.15)$, $P = 0.014$, $n = 31$ nests parasitized by day 4; slope for day 5 regression = $-0.29 (\pm 0.13)$, $P = 0.046$, $n = 45$ nests parasitized by day 5. I therefore chose day 3 as the cut-off point for separately examining the clutch-size responses of females who accepted or rejected parasitic eggs. For simplicity, I refer to these females as 'acceptors' and 'rejecters', respectively, but this grouping applies only to their one nest I examined and does not assume that they always reject or accept. In fact, egg rejection appears to be largely due to the chance difference between host and parasitic eggs.

The clutch-size comparisons provide indirect evidence for counting, but there is disagreement among cognitive psychologists over what the term 'counting' actually means^{1,24,26}. I adopt the terminology of Gallistel²⁴ who suggests that counting includes any "discriminations based on the numerosity of a set". Thus, 'counting' here indicates that birds are making decisions based on the number of eggs in their nests, not that they are performing addition or subtraction.

Received 12 November 2002; accepted 25 February 2003; doi:10.1038/nature01505.

1. Shettleworth, S. J. *Cognition, Evolution and Behavior* (Oxford Univ. Press, Oxford, 1998).
2. Rothstein, S. I. A model system for coevolution: avian brood parasitism. *Annu. Rev. Ecol. Syst.* **21**, 481–508 (1990).
3. Sherman, P. W., Reeve, H. K. & Pfennig, D. W. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 69–98 (Blackwell, Oxford, 1997).
4. Rothstein, S. I. Successes and failures in avian egg and nestling recognition with comments on the utility of optimality reasoning. *Am. Zool.* **22**, 547–560 (1982).
5. Davies, N. B., Brooke, M. de L. & Kacelnik, A. Recognition errors and probability of parasitism determine whether reed warblers should accept or reject mimetic cuckoo eggs. *Proc. R. Soc. Lond. B* **263**, 925–931 (1996).
6. Brooke, M. de L. & Davies, N. B. Egg mimicry by cuckoos *Cuculus canorus* in relation to discrimination by hosts. *Nature* **335**, 630–632 (1988).
7. Davies, N. B. *Cuckoos, Cowbirds and Other Cheats* (Poyser, London, 2000).
8. Rohwer, F. C. & Freeman, S. The distribution of conspecific nest parasitism in birds. *Can. J. Zool.* **67**, 239–253 (1989).
9. Eadie, J. M., Sherman, P. & Semel, B. in *Behavioral Ecology and Conservation Biology* (ed. Caro, T.) 306–340 (Oxford Univ. Press, Oxford, 1998).
10. Andersson, M. in *Producers and Scroungers* (ed. Barnard, C. J.) 195–228 (Croom Helm, London, 1984).
11. Lyon, B. E. *The Ecology and Evolution of Conspecific Brood Parasitism in American Coots* (Fulica americana) Thesis, Princeton Univ. (1992).
12. Lyon, B. E. Brood parasitism as a flexible female reproductive tactic in American coots. *Anim. Behav.* **46**, 911–928 (1993).
13. Lyon, B. E. Tactics of parasitic American coots: host choice and the pattern of egg dispersion among host nests. *Behav. Ecol. Sociobiol.* **33**, 87–100 (1993).
14. Lyon, B. E., Hochachka, W. M. & Eadie, J. M. Paternity-parasitism trade-offs: a model and test of host-parasite cooperation in an avian conspecific brood parasite. *Evolution* **56**, 1253–1266 (2002).
15. Kemeel, R. E. & Rothstein, S. I. Mechanisms of avian egg recognition: adaptive responses to eggs with broken shells. *Anim. Behav.* **36**, 175–183 (1988).
16. Jamieson, I. G., McRae, S. B., Simmons, R. E. & Trewby, M. High rates of conspecific brood parasitism

- and egg rejection in coots and moorhens in ephemeral wetlands in Namibia. *Auk* **117**, 250–255 (2000).
17. Arnold, T. W. Conspecific egg discrimination in American coots. *Condor* **89**, 675–676 (1987).
18. Jackson, W. M. in *Parasitic Birds and Their Hosts* (eds Rothstein, S. I. & Robinson, S. K.) 406–416 (Oxford Univ. Press, Oxford, 1998).
19. Bertram, B. C. R. *The Ostrich Communal Nesting System* 109–120 (Princeton Univ. Press, Princeton, 1992).
20. Davies, N. B. & Brooke, M. de L. An experimental study of co-evolution between the cuckoo, *Cuculus canorus*, and its hosts. II host egg markings, chick discrimination and general discussion. *J. Anim. Ecol.* **58**, 225–236 (1989).
21. Haywood, S. Sensory and hormonal control of clutch size in birds. *Q. Rev. Biol.* **68**, 33–60 (1993).
22. Andersson, M. & Eriksson, M. Nest parasitism in goldeneyes *Bucephala clangula*: some evolutionary aspects. *Am. Nat.* **120**, 1–16 (1982).
23. Lyon, B. E. Optimal clutch size and conspecific brood parasitism. *Nature* **392**, 380–383 (1998).
24. Gallistel, C. R. Counting versus subitizing versus the sense of number. *Behav. Brain Sci.* **11**, 585–586 (1988).

25. Hauser, M. D. What do animals think about numbers. *Am. Sci.* **88**, 144–151 (2000).
26. Seibt, U. Are animals naturally attuned to number? *Behav. Brain Sci.* **11**, 597–598 (1988).

Acknowledgements M. Andersson, J. Briskie, J. Eadie, R. Gallistel, G. Pogson, U. Seibt, S. Shettleworth, and in particular K. Wasson and A. Chaine provided comments on the manuscript. J. Estes, E. Geffen and C. Simms provided statistical advice. B. Bair, L. Cargill, S. Everding, L. Hamilton, D. Hansen, M. Magrath, and C. Morrill assisted in the field. P. Grant, R. Rubenstein and H. Horn provided advice during the study. The National Geographic Society, the National Science Foundation, the Chapman Fund and the Sigma Xi Society provided funding.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to B.L. (e-mail: lyon@biology.ucsc.edu).

An enhanced cosmic-ray flux towards ζ Persei inferred from a laboratory study of the $\text{H}_3^+ - \text{e}^-$ recombination rate

B. J. McCall^{*,†}, A. J. Huneycutt^{*}, R. J. Saykally^{*}, T. R. Geballe[‡], N. Djuric[§], G. H. Dunn[§], J. Semaniak^{||}, O. Novotny^{||}, A. Al-Khalili[#], A. Ehlerding[#], F. Hellberg[#], S. Kalhori[#], A. Neau[#], R. Thomas[#], F. Österdahl[☆] & M. Larsson[#]

^{*} Department of Chemistry, [†] Department of Astronomy, University of California at Berkeley, Berkeley, California 94720, USA

[‡] Gemini Observatory, 670 North A'ohoku Place, Hilo, Hawaii 96720, USA

[§] JILA, University of Colorado and National Institute of Standards and Technology, Boulder, Colorado 80309, USA

^{||} Institute of Physics, Świętokrzyska Academy, 25 406 Kielce, Poland

[#] Department of Electronics and Vacuum Physics, Faculty of Mathematics and Physics, Charles University Prague V Holesovickach 2, Prague 8, Czech Republic

[☆] Department of Physics, SCFAB, Stockholm University, S-106 91 Stockholm, Sweden

[☆] Manne Siegbahn Laboratory, Stockholm University, S-104 05 Stockholm, Sweden

The H_3^+ molecular ion plays a fundamental role in interstellar chemistry, as it initiates a network of chemical reactions that produce many molecules^{1,2}. In dense interstellar clouds, the H_3^+ abundance is understood using a simple chemical model, from which observations of H_3^+ yield valuable estimates of cloud path length, density and temperature^{3,4}. But observations of diffuse clouds have suggested that H_3^+ is considerably more abundant than expected from the chemical models^{5–7}. Models of diffuse clouds have, however, been hampered by the uncertain values of three key parameters: the rate of H_3^+ destruction by electrons (e^-), the electron fraction, and the cosmic-ray ionization rate. Here we report a direct experimental measurement of the H_3^+ destruction rate under nearly interstellar conditions. We also report the observation of H_3^+ in a diffuse cloud (towards ζ Persei) where the electron fraction is already known. From these, we find that the cosmic-ray ionization rate along this line of sight is 40 times faster than previously assumed. If such a high cosmic-ray flux is ubiquitous in diffuse clouds, the discrepancy between chemical models and the previous observations^{5–7} of H_3^+ can be resolved.

The dissociative recombination of H_3^+ (that is, the exothermic reaction $\text{H}_3^+ + \text{e}^- \rightarrow \text{H} + \text{H} + \text{H}$ or $\text{H}_2 + \text{H}$) has been one of the most controversial topics in the field of ion physics^{8,9}, as different experimental methods have disagreed by a factor of 10,000 on the value of the thermal rate coefficient. In addition, it was only in the past few years that a plausible theoretical mechanism (the “indirect process”¹⁰) was used to explain the dissociative recombination of H_3^+ in its lowest vibrational level¹¹. However, those calculations gave a cross-section¹² 1,000 times smaller than those measured by the ion storage rings CRYRING¹³ (Sweden), ASTRID¹⁴ (Denmark) and TARN II¹⁵ (Japan), which all gave essentially the same result. Calculations¹⁶ accounting for the symmetry-breaking of the equilateral-triangle-shaped H_3^+ ion by the incoming electron were in better accord with the experiments, and a more detailed theoretical treatment¹⁷ has now provided a step forward in terms of agreement, although important discrepancies with experiment remain to be resolved.

Meanwhile, a series of experiments at the Test Storage Ring in Germany^{18,19} and at CRYRING²⁰ suggested that, whereas H_3^+ undergoes complete vibrational relaxation by spontaneous emission of radiation in storage-ring experiments, the high rotational temperature of H_3^+ affected the measured values of the dissociative recom-

bination rate coefficient in the earlier experiments^{13–15}. These experiments^{18–20} thus made clear that a measurement of a thermal rate coefficient applicable to interstellar clouds would have to be performed on rotationally cold H_3^+ . This represented a challenge to the ion storage ring technique, which has traditionally used hot-filament or hollow-cathode ion sources that produce ions at much higher rotational temperatures than those of diffuse clouds.

The challenge of producing rotationally cold H_3^+ ions has been addressed by the development of a supersonic expansion ion source (see Fig. 1 top trace for results). The measured rate coefficient for the dissociative recombination of H_3^+ using this source is shown in Fig. 2. The general shape is similar to that reported in earlier studies^{13–15}, but the present data show a number of resonances between 1 meV and 1 eV. Although hints of resonance structures have been reported previously²⁰, the present experiment is, to our knowledge, the first to use rotationally cold ions, which apparently enhance the structure. Figure 3 shows the derived rate coefficient for a thermal electron distribution (k_e), which is about 40% lower than that measured in storage-ring studies of rotationally hot H_3^+ ions^{13,14}.

Whereas different ion storage ring experiments have consistently yielded about the same value of $k_e \approx 1 \times 10^{-7} \text{ cm}^3 \text{ s}^{-1}$ at 300 K (apart from the present experiment, which uses rotationally cold ions), afterglow experiments, which indirectly infer k_e from the removal of electrons in a decaying plasma of hydrogen and noble gases, are more difficult to interpret. Different afterglow measurements²¹ since 1984 have yielded values of k_e (at 300 K) which range from $<10^{-11}$ to $\sim 2 \times 10^{-7} \text{ cm}^3 \text{ s}^{-1}$. The reason for this wide

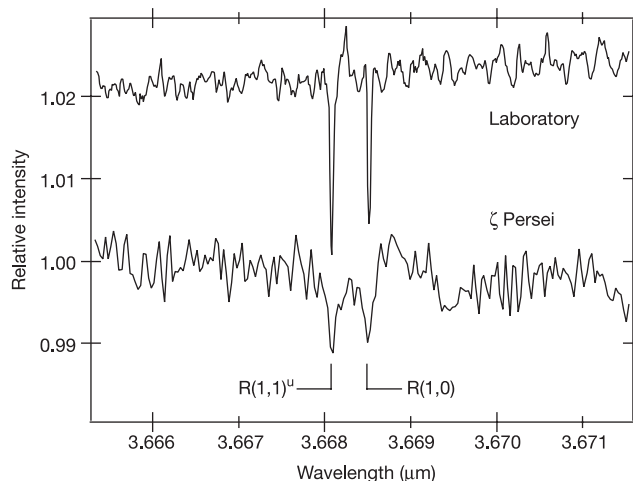


Figure 1 Spectra of two H_3^+ transitions arising from the two lowest rotational levels, which are the only levels with significant population in diffuse clouds. $\text{R}(1,1)^u$ originates from the lowest *para* level ($J = 1$, $K = 1$), while $\text{R}(1,0)$ comes from the lowest *ortho* level ($J = 1$, $K = 0$). Note that the ($J = K = 0$) level is forbidden by the Pauli principle. For energy-level diagrams and a description of the H_3^+ notation, see ref. 28. Top trace, cavity ringdown²⁹ laser absorption spectrum of the supersonic expansion ion source. In this source, H_3^+ number densities of $\sim 10^{11} \text{ cm}^{-3}$ were produced in a direct-current discharge plasma downstream of a 500- μm pinhole through which hydrogen gas (at 2.5 atm) expanded supersonically into a vacuum. Depending on conditions, the rotational temperature of the plasma was 20–60 K, on the basis of the relative intensities of the two transitions. This laboratory spectrum has been smoothed and multiplied by a factor of 1,000 for clarity. Bottom trace, spectrum of a diffuse cloud towards ζ Persei obtained using the CGS4 infrared spectrometer at the United Kingdom Infrared Telescope (UKIRT) on the night of 2001 September 5 UT using our standard observing techniques⁷. The total on-source integration time was 40 min, and the ζ Persei spectrum has been divided by that of a standard star (BS936) to remove absorption lines due to the Earth's atmosphere. The spectrum is displayed in rest wavelengths, and indicates an H_3^+ column density (the integral of the number density along the line of sight) of $N(\text{H}_3^+) = 8 \times 10^{13} \text{ cm}^{-2}$. The ratio of the two absorption lines yields a temperature estimate of 23 K.

variance in afterglow measurements is still not understood, although it may be related to the complicated modelling and analysis that goes into extracting the value of k_e from the experimental measurements. In contrast, the storage-ring experiments are conceptually very simple, and require a comparatively straightforward analysis to determine k_e , so we consider the present experiment to provide a much more robust measurement of the appropriate value of k_e for interstellar conditions. It is important to note two differences between the present experiment and interstellar conditions: the presence of high electron number densities ($n(e^-) \approx 10^7 \text{ cm}^{-3}$) and magnetic fields ($\sim 300 \text{ G}$). Because these are inherent to the storage-ring technique and cannot be avoided, our results need to be confirmed by theoretical calculations. Although there are still some unexplained differences between theory and our experiment, the most recent calculations¹⁷ yield a value of k_e ($\sim 2 \times 10^{-7} \text{ cm}^3 \text{ s}^{-1}$ at 40 K) that is similar to our results. Although we cannot explain the wide range in afterglow results, the weight of the evidence suggests that we have determined the appropriate value of k_e for interstellar conditions, and this fact compels us to examine the implications of adopting this new value for modelling H_3^+ chemistry in diffuse clouds.

The chemical model for H_3^+ in diffuse clouds is quite simple, in contrast to the involved models needed for most other molecules. H_3^+ is formed when H_2 is ionized by a cosmic ray (a high-energy proton or electron) to form H_2^+ , which then undergoes a very rapid ion–neutral reaction with another H_2 to form H_3^+ and an H atom.

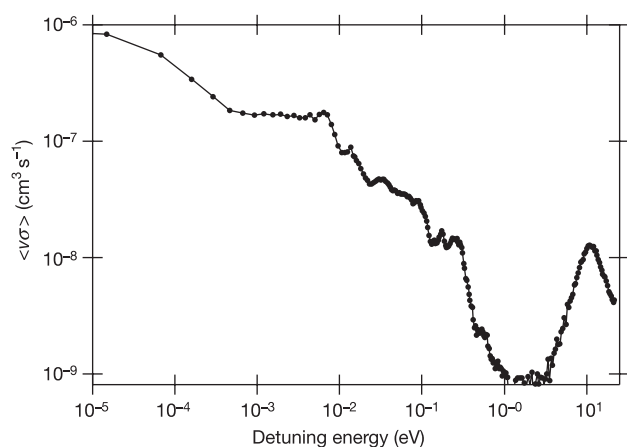


Figure 2 Measured dissociative recombination rate coefficient of rotationally cold H_3^+ as a function of detuning energy. For these measurements, the supersonic expansion ion source (described in Fig. 1 legend) was mounted onto the MINIS ion injection endstation of the CRYRING ion storage ring at the Manne Siegbahn Laboratory in Stockholm. The ions leaving the source were mass-selected and accelerated to 300 keV per a.m.u. before being injected into the ring. After injection, the ions were accelerated to their final energy (12.1 MeV), which gave a circulating ion beam current of 0.31 μA . In one straight section ('interaction region') of the 52-m-circumference storage ring, the ions passed collinearly through an approximately homogeneous electron beam of 85 cm length and 4 cm diameter ('electron cooler'). The electron beam current was 35.5 mA and the electron energy (2.2 keV) was chosen so that ion–electron velocity matching was achieved. For the first five seconds after the ion beam had reached its full energy, the electron beam acted to reduce the phase-space volume of the stored ion beam, leading to a narrow ion-velocity spread and a correspondingly small ion beam diameter ($\sim 1 \text{ mm}$). During this time, some stored ions were lost to dissociative recombination with electrons and collisions with residual gas molecules, while the remaining ions relaxed to the ground vibrational state by spontaneous emission. This cooling phase was followed by a measurement of the total number of neutral reaction products ($\text{H} + \text{H} + \text{H}$ and $\text{H}_2 + \text{H}$) using a surface barrier detector mounted tangentially to the ring following the interaction region. The electron energy was varied by changing the cathode voltage in the electron cooler, and the neutral fragment signal was recorded as a function of the longitudinal energy difference between electrons and ions ('detuning energy'). Note the resonance structure between 1 meV and 1 eV that was only hinted at in previous experiments using rotationally hot ions.

The rate of formation can be expressed as $\zeta n(\text{H}_2)$, where ζ is the cosmic-ray ionization rate and $n(\text{H}_2)$ is the number density of H_2 . The destruction of H_3^+ in diffuse clouds is dominated by dissociative recombination, and the rate of this process can be expressed as $k_e n(e^-) n(\text{H}_3^+)$. In steady state⁷, the H_3^+ number density can then be written as $n(\text{H}_3^+) = (\zeta/k_e) n(\text{H}_2)/n(e^-)$.

Astronomical absorption spectra, however, yield direct information only about the absorber's column density (the integral of the number density along the line of sight), not the number density itself. We assume that the ratio $n(\text{H}_2)/n(e^-)$ is approximately constant throughout the cloud, and replace it by the ratio of column densities $N(\text{H}_2)/N(e^-)$. Given this assumption, $n(\text{H}_3^+)$ is itself a constant, and we write $N(\text{H}_3^+) = n(\text{H}_3^+)L$, where L is the absorption path length. Substitution into the above equation then yields the relation $\zeta L = k_e [N(e^-)/N(\text{H}_2)] N(\text{H}_3^+)$.

Here we report the observation of H_3^+ in the diffuse cloud towards ζ Persei (Fig. 1 bottom trace); this is, to our knowledge, the first detection of H_3^+ along a line of sight where the electron fraction can be estimated. Assuming that most electrons come from the photoionization of carbon, we can use existing ultraviolet measurements of molecular hydrogen²² and ionized carbon²³ to derive the electron fraction $N(e^-)/N(\text{H}_2) = 3.8 \times 10^{-4}$. Combining this electron fraction, our observed H_3^+ column density, and the value $k_e = 2.6 \times 10^{-7} \text{ cm}^3 \text{ s}^{-1}$ from the CRYRING experiment (Fig. 3), we find $\zeta L \approx 8,000 \text{ cm}^2 \text{ s}^{-1}$ along this line of sight.

There is little direct observational information about the value of either ζ or L . It has generally been assumed that the value of ζ in diffuse clouds is the same as in dense clouds, where various measurements suggest a value of $\zeta = 3 \times 10^{-17} \text{ s}^{-1}$. The absorption path length can be indirectly estimated by dividing the total column density of hydrogen nuclei by an assumed number density. The total column density towards ζ Persei is $N_{\text{H}} \equiv N(\text{H}) + 2N(\text{H}_2) \approx 1.6 \times 10^{21} \text{ cm}^{-2}$ (adopting $N(\text{H}) = 6.3 \times 10^{20} \text{ cm}^{-2}$, ref. 24). The number density can be estimated from the rotational excitation of the CO and C_2 molecules, as well as from more detailed cloud models that consider the H/H_2 ratio²⁵. Although each of these methods has its uncertainties, they are all consistent with an average number density of $\langle n_{\text{H}} \rangle = 150\text{--}500 \text{ cm}^{-3}$. If we adopt a density of $\langle n_{\text{H}} \rangle = 250 \text{ cm}^{-3}$, we obtain a path length of $L = 2.1 \text{ pc}$.

Together, the 'canonical' value of ζ and the best estimate of L yield $\zeta L = 200 \text{ cm}^2 \text{ s}^{-1}$. However, the measured H_3^+ column density implies $\zeta L = 8,000 \text{ cm}^2 \text{ s}^{-1}$. If we adopt the canonical value of ζ ,

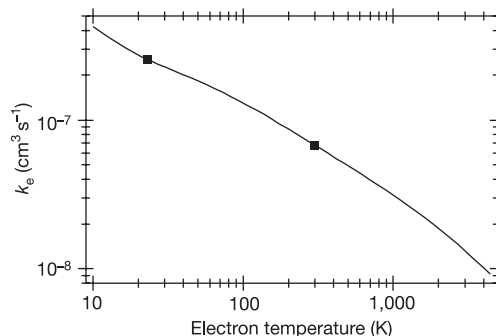


Figure 3 Calculated thermal rate coefficient k_e for the dissociative recombination of rotationally cold H_3^+ ions as a function of electron temperature, based on CRYRING measurements. The experimentally measured rate coefficient ($\langle \sigma \rangle$) shown in Fig. 2 was deconvolved using the known electron velocity distribution, yielding the dissociative recombination cross-section σ as a function of electron energy. The cross-section was then integrated over a Boltzmann distribution of energies (corresponding to a thermalized electron energy distribution) to yield the 'thermal' rate coefficient at each value of the electron temperature. The values $k_e(23 \text{ K}) = 2.6 \times 10^{-7}$ and $k_e(300 \text{ K}) = 6.8 \times 10^{-8} \text{ cm}^3 \text{ s}^{-1}$ are indicated with squares. Note that the value of $k_e(300 \text{ K})$ refers to an electron temperature of 300 K, but still pertains to rotationally cold H_3^+ ions. Details of the experimental method and analysis will be presented elsewhere.

we would have to accept a very long path length of $L = 85$ pc and a low density of $\langle n_{\text{H}} \rangle = 6 \text{ cm}^{-3}$. On the other hand, if we adopt the canonical path length, we must increase the ionization rate to $\zeta = 1.2 \times 10^{-15} \text{ s}^{-1}$. Given the evidence pointing to a short path length, and the fact that ζ is essentially unconstrained in diffuse clouds, we consider the higher ionization rate to be the more likely explanation. A higher value of ζ for diffuse clouds may also find some support from considerations of the abundance of the OH molecule²⁶. A more detailed discussion of the ζ Persei line of sight, as well as other observations of H_3^+ in diffuse clouds, will be presented elsewhere.

The high value of ζ suggested by the CRYRING measurements and the ζ Persei observations can be reconciled with the lower ionization rate in dense clouds by postulating the existence of a large flux of low-energy cosmic rays that can penetrate diffuse clouds but not dense clouds. Such a cosmic-ray flux would bring the models and observations of H_3^+ into agreement, but would also have far-reaching implications for the chemistry and physics of interstellar clouds. From a chemical perspective, a higher value of ζ would proportionately increase the number densities of oxygen compounds such as OH, as well as affecting the relative abundances of deuterium-bearing molecules. The high number density of H_3^+ in diffuse clouds would also increase the rate of proton-transfer reactions, which play the key role in the formation of complex molecules in dense clouds, but which have been considered relatively unimportant in diffuse clouds. From a physical perspective, a higher cosmic-ray ionization rate represents an additional heating source for interstellar gas, and could have a significant impact on the thermal balance of the warm neutral medium²⁷. Further observations of H_3^+ , H_2 and C^+ in lines of sight to other diffuse clouds are needed to determine if the ζ Persei line of sight is unusual, or whether a large, low-energy cosmic ray flux indeed pervades our galaxy. □

Received 19 November 2002; accepted 17 February 2003; doi:10.1038/nature01498.

- Herbst, E. & Klemperer, W. The formation and depletion of molecules in dense interstellar clouds. *Astrophys. J.* **185**, 505–534 (1973).
- Watson, W. D. The rate of formation of interstellar molecules by ion–molecule reactions. *Astrophys. J.* **183**, L17–L20 (1973).
- Geballe, T. R. & Oka, T. Detection of H_3^+ in interstellar space. *Nature* **384**, 334–335 (1996).
- McCall, B. J., Geballe, T. R., Hinkle, K. H. & Oka, T. Observations of H_3^+ in dense molecular clouds. *Astrophys. J.* **522**, 338–348 (1999).
- McCall, B. J., Geballe, T. R., Hinkle, K. H. & Oka, T. Detection of H_3^+ in the diffuse interstellar medium toward Cygnus OB2 No. 12. *Science* **279**, 1910–1913 (1998).
- Geballe, T. R., McCall, B. J., Hinkle, K. H. & Oka, T. Detection of H_3^+ in the diffuse interstellar medium: the Galactic center and Cygnus OB2 Number 12. *Astrophys. J.* **510**, 251–257 (1999).
- McCall, B. J. *et al.* Observations of H_3^+ in the diffuse interstellar medium. *Astrophys. J.* **567**, 391–406 (2002).
- Larsson, M. Experimental studies of the dissociative recombination of H_3^+ . *Phil. Trans. R. Soc. Lond. A* **358**, 2433–2444 (2000).
- Oka, T. Help!!! Theory for H_3^+ recombination badly needed. in *Dissociative Recombination of Molecular Ions with Electrons* (ed. Guberman, S. L.) (Kluwer Academic/Plenum, New York, in the press).
- Guberman, S. L. Dissociative recombination without a curve crossing. *Phys. Rev. A* **49**, R4277–R4280 (1994).
- Orel, A. E., Schneider, I. F. & Suzor-Weiner, A. Dissociative recombination of H_3^+ : progress in theory. *Phil. Trans. R. Soc. Lond. A* **358**, 2445–2456 (2000).
- Schneider, I. F., Orel, A. E. & Suzor-Weiner, A. Channel mixing effects in the dissociative recombination of H_3^+ with slow electrons. *Phys. Rev. Lett.* **85**, 3785–3788 (2000).
- Sundström, G. *et al.* Destruction rate of H_3^+ by low-energy electrons measured in a storage ring experiment. *Science* **263**, 785–787 (1994).
- Jensen, M. J. *et al.* Dissociative recombination and excitation of H_3^+ . *Phys. Rev. A* **63**, 052701 (2001).
- Tanabe, T. *et al.* in *Dissociative Recombination: Theory, Experiment and Applications* Vol. IV (eds Larsson, M., Mitchell, J. B. A. & Schneider, I. F.) 170–179 (World Scientific, Singapore, 2000).
- Kokouline, V., Greene, C. H. & Esry, B. D. Mechanism for the destruction of H_3^+ by electron impact. *Nature* **412**, 891–894 (2001).
- Kokouline, V. & Greene, C. H. Theory of dissociative recombination of D_3^+ triatomic ions, applied to H_3^+ . *Phys. Rev. Lett.* (in the press).
- Strasser, D. *et al.* Breakup dynamics and the isotope effect in H_3^+ and D_3^+ dissociative recombination. *Phys. Rev. A* **66**, 032719 (2002).
- Kreckel, H. *et al.* Vibrational and rotational cooling of H_3^+ . *Phys. Rev. A* **66**, 052509 (2002).
- Larsson, M. *et al.* Studies of dissociative recombination in CRYRING. in *Dissociative Recombination of Molecular Ions with Electrons* (ed. Guberman, S. L.) (Kluwer Academic/Plenum, New York, in the press).
- Plasil, R. *et al.* Advanced integrated stationary afterglow method for experimental study of recombination processes of H_3^+ and D_3^+ ions with electrons. *Int. J. Mass Spectrom.* **218**, 105–130 (2002).
- Savage, B. D., Drake, J. F., Budich, W. & Bohlin, R. C. A survey of interstellar molecular hydrogen. I. *Astrophys. J.* **216**, 291–307 (1977).

- Cardelli, J. A., Meyer, D. M., Jura, M. & Savage, B. D. The abundance of interstellar carbon. *Astrophys. J.* **467**, 334–340 (1996).
- Bohlin, R. C., Savage, B. D. & Drake, J. F. A survey of interstellar H I from $\text{Ly}\alpha$ absorption measurements. II. *Astrophys. J.* **224**, 132–142 (1978).
- van Dishoeck, E. F. & Black, J. H. Comprehensive models of diffuse interstellar clouds — physical conditions and molecular abundances. *Astrophys. J. Suppl. Ser.* **62**, 109–145 (1986).
- Lepp, S. in *Astrochemistry of Cosmic Phenomena* (ed. Singh, P. D.) 471–475 (IAU Symposium 150, Kluwer, Dordrecht, 1992).
- Heiles, C. & Troland, T. H. The millennium Arecibo 21-cm absorption line survey. II. Properties of the warm and cold neutral media. *Astrophys. J.* (in the press); preprint astro-ph/0207105 at (<http://xxx.lanl.gov>) (2002).
- Lindsay, C. M. & McCall, B. J. Comprehensive evaluation and compilation of H_3^+ spectroscopy. *J. Mol. Spectrosc.* **210**, 60–83 (2001).
- Paul, J. B., Collier, C. P., Saykally, R. J., Scherer, J. J. & O’Keefe, A. Direct measurement of water cluster concentrations by infrared cavity ringdown laser absorption spectroscopy. *J. Phys. Chem. A* **101**, 5211–5214 (1997).

Acknowledgements B.J.M. and A.J.H. thank K. Wilson and C.-Y. Chung for their assistance in the laboratory, D. Lucas for the loan of equipment, and H. Chan and E. Granlund for their support. B.J.M. thanks T. Oka, L. M. Hobbs, D. G. York and T. P. Snow for conversations about the ζ Persei line of sight. We thank C. H. Greene for providing us with the results of his calculations in advance of publication. M.L. thanks D. Zajfman for information on TSR results before publication. The idea of using a supersonic expansion source for dissociative recombination measurements originated during conversations between B.J.M. and C. M. Lindsay, and the experiment was designed by B.J.M., A.J.H. and M.L. The supersonic expansion ion source was built and spectroscopically characterized by A.J.H. and B.J.M. in the laboratory of R.J.S. in Berkeley. The dissociative recombination measurements in Stockholm were carried out by all of the authors except R.J.S. and T.R.G. The UKIRT observations were obtained by B.J.M. and T.R.G. The authors thank the staff of the Manne Siegbahn Laboratory for help with the experiment. B.J.M. is supported by the Miller Institute for Basic Research in Science. A.J.H. and R.J.S. acknowledge support from the AFOSR and NSF. T.R.G. is supported by Gemini Observatory, operated by AURA for the international Gemini partnership. N.D. and G.H. were supported in part by the US DOE, Office of Fusion Energy. J.S. is supported in part by the State Committee for Scientific Research. Support is acknowledged from the EU fifth framework program, the EOARD, the Swedish Research Council and STINT.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.J.M. (e-mail: bjmcalls@astro.berkeley.edu).

Structure and thermal history of the H-chondrite parent asteroid revealed by thermochronometry

Mario Trieloff*, Elmar K. Jessberger†, Ingrid Herrwerth‡, Jens Hopp*, Christine Fiéni§, Marianne Ghéllis§, Michèle Bourot-Denise§ & Paul Pellas§||

* Mineralogisches Institut der Universität Heidelberg, Im Neuenheimer Feld 236, D-69120 Heidelberg, Germany

† Institut für Planetologie, Westfälische Wilhelms-Universität, Wilhelm-Klemm-Straße 10, D-48149 Münster, Germany

‡ Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, D-69117 Heidelberg, Germany

§ Laboratoire de Minéralogie du Muséum, 61 rue Buffon, 75005 Paris, France

|| Deceased

Our Solar System formed ~4.6 billion years ago from the collapse of a dense core inside an interstellar molecular cloud. The subsequent formation of solid bodies took place rapidly. The period of <10 million years over which planetesimals were assembled can be investigated through the study of meteorites^{1–3}. Although some planetesimals differentiated and formed metallic cores like the larger terrestrial planets, the parent bodies of undifferentiated chondritic meteorites experienced comparatively mild thermal metamorphism that was insufficient to separate metal from silicate^{4,5}. There is debate about the nature of the heat source^{6–9} as well as the structure and cooling

history of the parent bodies^{10–12}. Here we report a study of ²⁴⁴Pu fission-track and ⁴⁰Ar–³⁹Ar thermochronologies of unshocked H chondrites, which are presumed to have a common, single, parent body. We show that, after fast accretion, an internal heating source (most probably ²⁶Al decay^{8–10,13}) resulted in a layered parent body⁶ that cooled relatively undisturbed: rocks in the outer shells reached lower maximum metamorphic temperatures and cooled faster than the more recrystallized and chemically equilibrated rocks from the centre, which needed ~160 Myr to reach 390 K.

The first condensation of refractory Ca–Al-rich inclusions (Allende CAIs) in the solar nebula occurred $4,566 \pm 2$ Myr ago^{1,2}. Accretion and differentiation into silicate crust and metallic core of some asteroids was completed only a few million years later, as demonstrated by Pb–Pb ages of basaltic achondrites³, whereas cooling of the iron cores extended over tens of millions of years¹⁴. In contrast, the cooling history and structure of the parent bodies of undifferentiated chondrites are not well established. Ordinary chondrites of one class (H, L or LL) share common properties such as oxygen isotopic composition, degree of iron oxidation and major element chemistry, suggesting a single parent body for each class. Moreover, each class can be subdivided into different petrologic types^{4,5} that experienced specific degrees of thermal metamorphism: for example, type 6 chondrites were heated to higher metamorphic peak temperatures (~1,120 K) than those of type 4 (~920 K). A possible model for the early evolution of ordinary chondrite parent asteroids is an onion-shell-layered parent body (refs 6, 10) (Fig. 1). After accretion, the parent body consists of primitive material and is heated internally by decay energy of short-lived nuclides^{8,9,13} such as ²⁶Al ($\tau_{1/2} = 0.72$ Myr). The outer shells lose heat more effectively, whereas the rocks in the centre are heated to higher metamorphic peak temperatures and cool down more slowly. However, as some studies did not show the expected correlation between metamorphic grade and cooling rate or age^{15,16}, other hypotheses were advanced that involve either separate parent bodies for each petrologic type¹¹ or early collisional fragmentation of an onion-shell-layered parent body, with final cooling in a ‘rubble-pile’ structure¹². Moreover, alternative heat sources were considered (see review⁷), such as electric currents induced by an early, intense, solar-wind plasma.

Radiochronometry can elucidate the early parent asteroid evolution: in an onion-shell-layered parent body that cools undis-

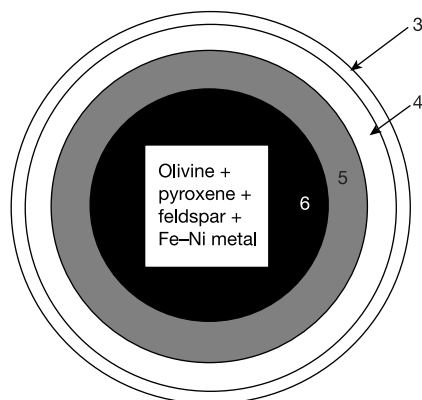


Figure 1 Onion-shell model for undifferentiated ordinary chondrite parent asteroids⁶. A celestial body that is internally heated (probably by the decay energy from short-lived ²⁶Al) reaches higher maximum metamorphic temperatures in its centre, resulting in higher petrologic types (numbers shown) that cool slower than in the outer layers, which in turn results in lower petrologic types with faster cooling. Increasing degrees of thermal metamorphism resulted, for example, in the extinction of primary chondrules, matrix recrystallization, formation of secondary feldspar and phosphates, and chemical equilibration of olivine and pyroxene. However, heating was insufficient for separation of metal and silicate and formation of an iron core.

turbed, type 6 chondrites cooled slower than type 4 chondrites. Slower cooling should result in lower cooling ages, as radiometric ‘ages’ actually represent cooling below a ‘closure temperature’ that is mineral specific, for example when Pb diffusion in phosphates becomes ineffective below ~720 K (refs 17–19), or when ⁴⁰Ar from ⁴⁰K decay is retained quantitatively in oligoclase feldspar below ~550 K (refs 19, 20). Other examples are ²⁴⁴Pu-derived fission tracks stable in orthopyroxene below ~550 K (ref. 19) or in the phosphate merrillite below ~390 K (ref. 19). Most reliable

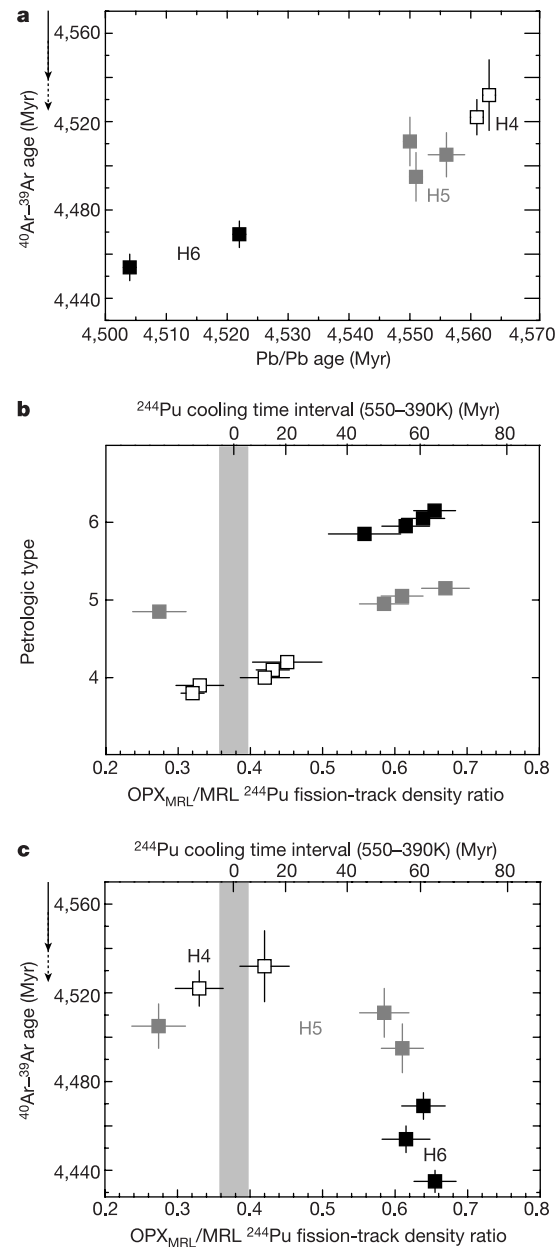


Figure 2 Correlation between cooling ages and metamorphic grade. **a**, Pb/Pb phosphate and ⁴⁰Ar–³⁹Ar feldspar cooling ages demonstrate that low petrographic types cooled faster to Pb closure in phosphates (~720 K) and Ar retention in feldspar (~550 K) than higher petrographic types. The arrow quantitatively indicates a possible shift of the K–Ar timescale caused by systematic error of the K decay constant, but note that age differences are not affected. **b**, Low petrographic types also cooled faster between 550 K and 390 K, as derived from cooling time intervals (upper scale) calculated from the ²⁴⁴Pu fission-track density ratio (lower scale) in merrillite and adjacent orthopyroxene (see Supplementary Table S1). The systematic uncertainty of the ²⁴⁴Pu timescale (induced by the normalization factor) is quantitatively shown by the grey band around zero. **c**, Correlation between ⁴⁰Ar–³⁹Ar feldspar cooling age and ²⁴⁴Pu fission-track cooling time interval.

results are obtained when the different complementary methods (Pb–Pb, ^{40}Ar – ^{39}Ar , ^{244}Pu fission tracks) are simultaneously applied to the aforementioned minerals that are major carriers of the parent nuclide. However, owing to the fine-grained nature of chondritic material, ^{40}Ar – ^{39}Ar dating is commonly done on whole rocks^{19,20}. Here we present the first combined feldspar, pyroxene and whole-rock ^{40}Ar – ^{39}Ar data for ordinary chondrites (see Supplementary Figs S1–S6 and Table S2). We also present the first ^{244}Pu fission-track chronometric results that were evaluated with all-recent methodological improvements^{19,21} (Supplementary Table S1). An indispensable requirement for obtaining information on the early parent-body history is to study exceptionally undisturbed chondrites. Only these preserve the earliest radiometric record from the period shortly after accretion, as they were not affected by secondary-impact-induced shock or reheating during the past ~4,000 Myr. Just a few unshocked H chondrites—like those we studied—satisfy this condition.

Table 1 provides a summary of ^{244}Pu fission-track, ^{40}Ar – ^{39}Ar , and Pb–Pb chronometries of all H chondrites we studied, for which complete data sets are available now. The highest ages are Pb–Pb phosphate ages of H4 chondrites Forest Vale and Ste Marguerite. Obviously, accretion was complete 3.0 ± 2.6 Myr after Allende CAI formation 4,566 Myr ago (refs 1, 2, 18). H5 and H6 chondrites cooled progressively later below the Pb–Pb closure temperature. Our new results extend this trend to low-temperature chronometers: individual ^{40}Ar – ^{39}Ar ages are generally lower than the respective Pb–Pb age (Table 1 and Fig. 2a). This can be interpreted as additional cooling time needed to reach ^{40}Ar retention in oligoclase feldspar at ~550 K. Moreover, high petrologic types needed a longer cooling time than low petrologic types, and reached isotopic closure later, leading to younger apparent ages. Slower cooling of H chondrites of high petrologic type is also confirmed by the cooling time intervals calculated from ^{244}Pu fission-track densities of orthopyroxene and merrillite, which register tracks at <550 and <390 K (Table 1). This correlation is also valid for other unshocked H chondrites for which, however, there are no complementary ^{40}Ar – ^{39}Ar and Pb–Pb age data (Fig. 2b and Supplementary Table S1). Figure 2c shows that H5 and H6 chondrites, which needed longer for cooling to the Ar closure temperature of 550 K,

also cooled more slowly afterwards until ^{244}Pu fission-track retention in merrillite at 390 K (except Nadiabondi).

Figure 3a displays integrated cooling curves: following the accretion shortly after 4,566 Myr (refs 1, 2), rapid heating resulted in different metamorphic peak temperatures, between 920 K and 1,120 K (ref. 5), at different depths. H4 Forest Vale and Ste Marguerite cooled down to the Pb closure temperature at 720 K within a few million years (4,561 and 4,563 Myr ago), whereas types H5 needed 10–16 Myr, and H6 Kernouve and Guarena considerably longer¹⁸. Subsequent cooling until ^{40}Ar retention in oligoclase at ~550 K needed additional time, and occurred later for H6 than for H5 and for H4. As ^{244}Pu fission tracks in orthopyroxene are also retained at ~550 K, ^{244}Pu cooling-time intervals until track registration in merrillite at ~390 K can thus be attached to the ^{40}Ar – ^{39}Ar data: consequently H4 chondrites cooled to 390 K faster and earlier than H5 and H6 chondrites. Note that H6 Guarena, H6 Kernouve and H6 Estacado cooled slowly, but with significant differences, whereas the H5 Allegan and H5 Richardton have rather similar cooling histories. H4 Forest Vale and Ste Marguerite are also very similar, whereas H5 Nadiabondi seems to be intermediate between H5 and H4, as indicated by its Pb–Pb age and its steep ^{244}Pu fission-track cooling rate. Apparently, this H5 and the two H4 chondrites share a rather peculiar cooling history: fast cooling above 720 K and below 550 K seems to be interrupted by a slower cooling episode between 720 K and 550 K. However, this is most probably not significant, taking into account the recently suggested revision of the relative Pb–Pb and K–Ar age scales^{22–25}, resulting from the uncertainty in the K decay constant. A shift of the ^{40}Ar – ^{39}Ar ages (and also ^{244}Pu ages) by 30 Myr (ref. 25) obviously provides a more concise explanation (Fig. 3b): then Ste Marguerite, Forest Vale and Nadiabondi cooled fast through the high- and low-temperature regimes, instead of having rather exotic ‘fast–slow–fast’ cooling histories. It is important to note that the K–Ar and ^{244}Pu age scales are shifted to the same degree, because the fission-track retention in orthopyroxene is linked to ^{40}Ar retention in feldspar. Although our data support the need for a revision of the K–Ar age scale, its exact magnitude^{23–25} remains an issue.

We also present in Fig. 3b the modelled cooling curves at different depths of a chondritic asteroid 100 km in diameter that was heated

Table 1 Summary of thermochronological data

Meteorite	^{244}Pu fission track density ratio of orthopyroxene (merrillite contacts) and merrillite*	^{244}Pu cooling time interval† (Myr)	^{40}Ar – ^{39}Ar age‡ relative to NL25 homblende standard§ (Myr)	Pb–Pb age (Myr)
Retention/closure temperature		550K (Opx), 390K (Mrl)	550 ± 20K (Oligoclase feldspar)	720 ± 50K (Phosphates)
H4				
Ste Marguerite	0.420 ± 0.034	12 ± 11	4,532 ± 16 WR	4,563 ± 1
Forest Vale	0.330 ± 0.033	<8¶	4,522 ± 8 WR	4,561 ± 1
H5				
Nadiabondi	0.261 ± 0.017	<–8#	4,505 ± 10 WR, Px	4,556 ± 3
Richardton	0.610 ± 0.029	56 ± 9	4,495 ± 11 WR	4,551 ± 1
Allegan	0.585 ± 0.034	51 ± 9	4,511 ± 11 WR	4,550 ± 1
Sena	0.670 ± 0.033	66 ± 9		4,556 ± 1
H6				
Kernouve	0.639 ± 0.030	61 ± 8	4,469 ± 6 WR, Fs, Px	4,522 ± 1
Guarena	0.615 ± 0.033	56 ± 9	4,454 ± 6 WR, Fs, Px	4,504 ± 1
Estacado	0.655 ± 0.029	64 ± 8	4,435 ± 5 WR, Fs, Px	

* ^{244}Pu fission-track densities in orthopyroxene (Opx) crystal surfaces that had been located adjacent to merrillite (Mrl) phosphate crystals; merrillite crystals were pre-annealed at 290 °C to erase cosmic-ray-induced spallation recoil tracks¹⁹.

† The cooling time interval is obtained after correcting the fission-track densities for irradiation geometry and mineral registration efficiency. This correction factor induces an additional systematic uncertainty of the ^{244}Pu timescale, which need not be taken into account for comparison of relative cooling interval differences (see Methods).

‡ ^{40}Ar – ^{39}Ar ages are error-weighted means of plateau ages of whole rock (WR), feldspar (Fs) and/or pyroxene (Px) separates for H6 chondrites. Plateaux were also obtained for Nadiabondi WR and Forest Vale WR. For Richardton, Allegan and Ste Marguerite we calculated total ages (see Supplementary Figs S2 and S3). Note that in all chondrites feldspar is the dominant K-bearing and, hence, actually dated phase (see Methods and Supplementary Figs S1–S4). There is a very good agreement of ^{40}Ar – ^{39}Ar ages obtained²⁰ on four of these H chondrites by using a different age monitor (see Supplementary Table S2).

§ Individual K–Ar ages were calculated by using the Steiger and Jäger²² convention, and are calibrated against the age-standard NL25 homblende²⁰ with a K–Ar age of $2,655 \pm 9$ Myr (see Supplementary Table S2). Uncertainties do not contain the systematic error due to the NL25-standard age error (a possible ± 12 Myr shift of the ^{40}Ar – ^{39}Ar age scale) and the K-decay constant uncertainty (an upward shift of ages of the order of 30–45 Myr)^{23–25}.

|| Reference 18.

¶ Upper limit for cooling time interval at 2σ level.

Negative cooling time interval within 2σ uncertainties indicates fast cooling.

by ^{26}Al decay (with $^{26}\text{Al}/^{27}\text{Al} = 4 \times 10^{-6}$) 2.5 Myr after Allende CAI formation ($^{26}\text{Al}/^{27}\text{Al} = 5 \times 10^{-5}$) (ref. 8). We used the analytical model of Miyamoto *et al.*¹⁰ (see Methods). Typical cooling curves provide good approximations to the measured cooling histories, with the H6 Estacado, H6 Guarena and H6 Kernouve being deepest and H4 Forest Vale and H4 Ste Marguerite (and possibly Nadiabondi) lying shallowest. It is apparent that this rather simple model not only provides general features like slower cooling in the centre and faster cooling near the surface, but also reproduces the non-exponential cooling paths of the H5 chondrites. Nevertheless, more refined models (for example, those that take into account effects of the slow development of insulating regolith layers that allow faster cooling in the low-temperature regime between 550 K and 390 K for the H4 chondrites) could probably provide even better fits to the data.

Our new results define the thermal history of the H-chondrite parent body to a much greater precision than previous studies not using rigorously unshocked or undisturbed samples¹⁵ or mineral separates¹⁶. It agrees with the thermal history from I-Xe studies of chondritic mineral separates (ref. 26). It supports an internal heating mechanism, and favours decay energy of short-lived nuclides (for example ^{26}Al) as the planetesimal heat source in the early solar system^{7,9,13}. Obviously, the interior regions cooled undisturbed down to 390 K, indicating that this parent body was

not destroyed by early impacts. Fragmentation and re-assembly processes causing a rubble-pile structure of asteroids as currently observed²⁷ probably occurred much later. The early histories of other ordinary chondrite parent bodies remain to be assessed, but complementary radiometric data on mineral separates of unshocked L and LL chondrites with comparable precision are not yet available^{11,15,18,20,26}. □

Methods

Sample selection

Investigating the earliest parent-body history 4,600 Myr ago requires samples that were unaffected by impact-induced shock or heating during the past 4,000 Myr. Thus we applied three criteria to select our samples: (1) absence of shock features (see Supplementary Table S1); (2) presence of ^{244}Pu fission tracks—short-lived ^{244}Pu produces detectable tracks only >4,000 Myr ago, and subsequent reheating to >390 K is excluded because annealing temperatures (in merrillite) are low; (3) absence of major disturbances of the Pb–Pb system in phosphates. Additional studies of even only moderately shocked H4 chondrites (Ochansk, Menow; Supplementary Fig. S3c) revealed irregularities and inconclusive chronological results that simulate complex parent-body histories.

^{244}Pu fission-track thermochronometry

Short-lived ^{244}Pu ($\tau_{1/2} = 80$ Myr) is concentrated in the phosphates apatite and merrillite. Spontaneous ^{244}Pu fission leaves radiation damage visible by etching as fission tracks. These are also registered by accidentally adjacent silicate minerals²⁸. Upon cooling below a certain temperature, radiation damage is no longer annealed and fission tracks are retained, for example in orthopyroxene below ~550 K and in merrillite below ~390 K (ref. 19). If—appropriately corrected¹⁹—the fission-track densities in both minerals are the same, ‘instantaneous’ fast cooling is indicated, whereas a fission-track density ratio of 2 corresponds to a cooling time interval of 80 Myr at an average rate of 2 K Myr^{-1} . Important major corrections to fission-track dating introduced in recent years are (1) erasure of cosmic-ray-induced spallation recoil tracks in phosphates by crystal tempering at 290 °C (refs 19, 21), and (2) correction for different mineral-specific registration efficiencies. The more effective track registration in merrillite compared with that in orthopyroxene results in an effective correction factor of 1.325 ± 0.074 (average of previously used values^{19,21}) for the orthopyroxene/merrillite fission-track density ratio. To account for the irradiation geometry an additional factor of 2 has to be included. To check for relative cooling histories of meteorites, that is, cooling time interval differences, it is, however, appropriate to compare the uncorrected ratios (Table 1 and Fig. 2b, c). This is because the systematic uncertainties of the registration efficiency factor need not be included, which is the case when calculating absolute ^{244}Pu cooling-time intervals (Table 1 and Fig. 2b, c).

^{40}Ar – ^{39}Ar cooling ages

In ^{40}Ar – ^{39}Ar dating we studied whole rock samples, but also feldspar and pyroxene mineral separates of H6 chondrites, where thermal metamorphism had led to growth of large (>50 μm) grains of secondary feldspar^{4,5}. All 15 samples were analysed with high resolution (up to 42 heating steps), crucial for obtaining the spectral fine structure. The spectra (Table 1, and Supplementary Table S2 and Figs S1–S3) exclude significant secondary ^{40}Ar losses. However, those of the low petrologic-type chondrites are somewhat disturbed by ^{39}Ar recoil as expected for fine-grained mineral assemblages. The fine structures of such age spectra that are disturbed by ^{39}Ar recoil, but which still exhibit intermediate age plateaux, have previously been modelled quantitatively²⁹. It was demonstrated that the plateau segments yield reliable age information whereas the ages from initial and final extractions are to be discarded. As is evident from the K/Ca ratio spectra and the release patterns of K-derived ^{39}Ar and Ca-derived ^{37}Ar (Supplementary Figs S1–S4), the chronological information of whole rock samples is controlled by the main K carrier, oligoclase feldspar. Even in pyroxene mineral separates analysed by us, most K was located in fine-grained feldspar impurities that were not successfully separated (Supplementary Figs S1, S2 and S4). This is consistent with our observation of no significant age difference between whole rock, feldspar and pyroxene separates (Table 1 and Supplementary Figs S1 and S2 and Table S2). Accordingly, the average ^{40}Ar – ^{39}Ar age calculated from these different plateau ages from a single H chondrite can be directly interpreted as the feldspar cooling age. To minimize relative uncertainties, all samples were calibrated against the same NL25 age standard³⁰. Ages were calculated by using the convention of Steiger and Jäger²², but a recalibration of the relative K–Ar and Pb–Pb age scales should be considered^{23–25} (see main text). For irradiation details, see Supplementary Table S2 and Fig. S6.

Parent body heating model

In the calculation of the cooling curves of material in a chondritic parent body that is heated by the decay of ^{26}Al (Fig. 2b), we applied the analytical model of Miyamoto *et al.*¹⁰. As in ref. 10 our parameters are thermal conductivity (1.0 W m K^{-1}), density (3.2 g cm^{-3}), heat generation ($11.67 \times ^{26}\text{Al}/^{27}\text{Al} \text{ W m}^{-3}$) and emissivity. We slightly adapted the radius (100 instead of 85 km), accretion temperature T_0 (300 instead of 200 K), and initial $^{26}\text{Al}/^{27}\text{Al}$ ratio (4×10^{-6} instead of 5×10^{-6}). This $^{26}\text{Al}/^{27}\text{Al}$ ratio is a factor of 12.5 lower than in CAIs and corresponds to H chondrite accretion 2.5 Myr after the formation of Allende CAIs. It is in agreement with the Pb–Pb age difference of CAIs of 4,566 Myr and Ste Marguerite of 4,563 Myr (refs 1, 2, 18).

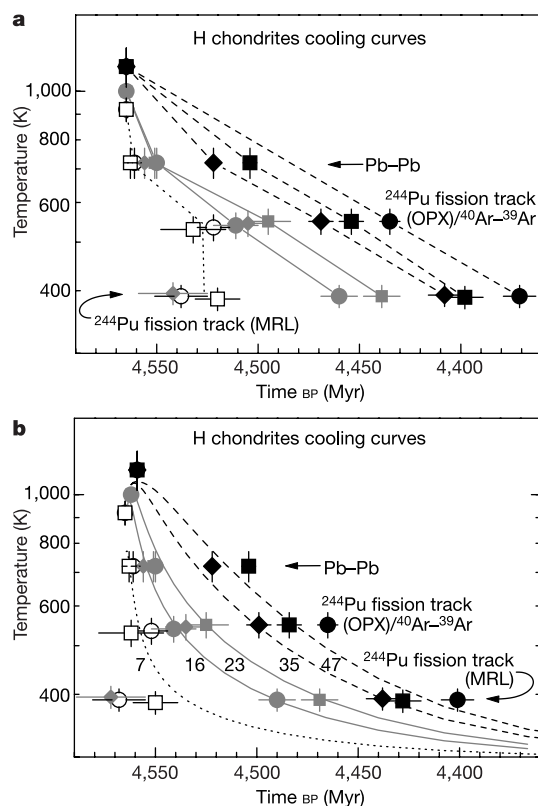


Figure 3 Integrated cooling curves for all H chondrites where complete chronological information is available. The Pb–Pb (at 720 K) and ^{40}Ar – ^{39}Ar (at 550 K) cooling ages and ^{244}Pu fission track retention ages in orthopyroxene (OPX, at 550 K) and merrillite (MRL, at 390 K) are given. Low petrographic types cooled significantly faster than higher ones. **a**, ^{40}Ar – ^{39}Ar ages as calculated after Steiger and Jäger²². Dashed lines: H6 Estacado (filled circles), Guarena (filled square), Kernouve (filled diamond). Solid lines: H5 Richardton (grey square), Allegan (grey circle), Nadiabondi (grey diamond). Dotted lines: H4 Forest Vale (open circle), Ste Marguerite (open square). **b**, Alternative hypothesis resulting from shifting ^{40}Ar – ^{39}Ar and ^{244}Pu ages up by 30 Myr to account for recent suggestions to recalibrate the Pb/Pb and Ar–Ar age scales^{23–25}. The cooling curves are calculated for a chondritic 100-km diameter asteroid that is internally heated by ^{26}Al decay for depths of 47, 35 (dashed), 23, 16 (solid) and 7 km (dotted). Symbols as in **a**.

Received 13 December 2002; accepted 18 February 2003; doi:10.1038/nature01499.

- Chen, J. H. & Wasserburg, G. J. The isotopic composition of uranium and lead in Allende inclusions and meteoritic phosphates. *Earth Planet. Sci. Lett.* **5**, 15–21 (1981).
- Allègre, C. J., Manhès, G. & Göpel, C. The age of the Earth. *Geochim. Cosmochim. Acta* **59**, 1445–1456 (1995).
- Lugmair, G. W. & Galer, S. J. G. Age and isotopic relationships among the angrites Lewis Cliff 86010 and Angra dos Reis. *Geochim. Cosmochim. Acta* **56**, 1673–1694 (1992).
- van Schmus, W. R. & Wood, J. A. A chemical-petrologic classification for the chondritic meteorites. *Geochim. Cosmochim. Acta* **31**, 747–765 (1967).
- Dodd, R. T. Metamorphism of the ordinary chondrites: a review. *Geochim. Cosmochim. Acta* **33**, 161–203 (1969).
- Herndon, J. M. & Herndon, M. A. Aluminium-26 as a planetoid heat source in the early solar system. *Meteoritics* **12**, 459–465 (1977).
- Wood, J. A. & Pellas, P. in *The Sun in Time* (eds Sonett, C. P., Giampapa, M. S. & Matthews, M. S.) 741–760 (Univ. of Arizona Press, Tucson, 1991).
- Lee, T., Papanastassiou, D. A. & Wasserburg, G. W. Demonstration of ^{26}Mg excess in Allende and evidence for ^{26}Al . *Geophys. Res. Lett.* **3**, 109–112 (1976).
- Knödseder, J., Cervino, M., Le Duigou, J. M., Meynet, G., Schaefer, D. & von Ballmoos, P. Gamma-ray line emission from OB associations and young open clusters—II. The Cygnus region. *Astron. Astrophys.* **390**, 945–960 (2002).
- Miyamoto, M., Fujii, N. & Takeda, H. Ordinary chondrite parent body: an internal heating model. In *Proc. Lunar Planet. Sci. Conf.* **12B**, 1145–1152 (1981).
- Scott, E. R. D. & Rajan, R. S. Metallic minerals, thermal histories and parent bodies of some xenolithic, ordinary chondrite meteorites. *Geochim. Cosmochim. Acta* **45**, 53–67 (1981).
- Grimm, R. E. Penecontemporaneous metamorphism, fragmentation, and reassembly of ordinary chondrite parent bodies. *J. Geophys. Res.* **90**(B2), 2022–2028 (1985).
- Zinner, E. & Göpel, C. Aluminium-26 in H4 chondrites: implications for its production and its usefulness as a fine-scale chronometer for early solar system events. *Meteoritics Planet. Sci.* **37**, 1001–1013 (2002).
- Wood, J. A. The cooling rates and parent planets of several iron meteorites. *Icarus* **3**, 429–459 (1964).
- Taylor, G. J., Maggiori, P., Scott, E. R. D., Rubin, A. E. & Keil, K. Original structures, and fragmentation and reassembly histories of asteroids: evidence from meteorites. *Icarus* **69**, 1–13 (1987).
- Jordan, J., Kirsten, T. & Richter, H. $^{129}\text{I}/^{127}\text{I}$: a puzzling early solar system chronometer. *Z. Naturforsch.* **35**(a), 145–170 (1980).
- Cherniak, D., Lanford, W. A. & Ryerson, F. J. Lead diffusion in apatite and zircon using ion implantation and Rutherford backscattering techniques. *Geochim. Cosmochim. Acta* **55**, 1663–1673 (1991).
- Göpel, C., Manhès, G. & Allègre, C. J. U–Pb systematics of phosphates from equilibrated ordinary chondrites. *Earth Planet. Sci. Lett.* **121**, 153–171 (1994).
- Pellas, P., Fieni, C., Trieloff, M. & Jessberger, E. K. The cooling history of the Acapulco meteorite as recorded by the ^{244}Pu and ^{40}Ar – ^{39}Ar chronometers. *Geochim. Cosmochim. Acta* **61**, 3477–3501 (1997).
- Turner, G., Enright, M. C. & Cadogan, P. H. The early history of chondrite parent bodies inferred from ^{40}Ar – ^{39}Ar ages. *Proc. Lunar Planet. Sci. Conf.* **9**, 989–1025 (1978).
- Lavielle, B., Marti, K., Pellas, P. & Perron, C. Search for extinct ^{248}Cm in the early Solar System. *Meteoritics* **27**, 382–386 (1992).
- Steiger, R. H. & Jäger, E. Subcommission on Geochronology: convention on the use of decay constants in geo- and cosmochronology. *Earth Planet. Sci. Lett.* **36**, 359–362 (1977).
- Renne, P. R. ^{40}Ar – ^{39}Ar age of plagioclase from Acapulco meteorite and the problem of systematic errors in cosmochronology. *Earth Planet. Sci. Lett.* **175**, 13–26 (2000).
- Begemann, F. et al. Call for an improved set of decay constants for geochronological use. *Geochim. Cosmochim. Acta* **65**, 111–121 (2001).
- Trieloff, M., Jessberger, E. K. & Fieni, C. Comment on ‘ ^{40}Ar – ^{39}Ar age of plagioclase from Acapulco meteorite and the problem of systematic errors in cosmochronology’ by Paul R. Renne. *Earth Planet. Sci. Lett.* **190**, 267–269 (2001).
- Brazzle, R. H., Pravdivtseva, O. V., Meshik, A. P. & Hohenberg, C. M. Verification and interpretation of the I–Xe chronometer. *Geochim. Cosmochim. Acta* **63**, 739–760 (1999).
- Pravec, P. & Harris, A. W. Fast and slow rotation of asteroids. *Icarus* **148**, 12–20 (2000).
- Pellas, P. & Storzer, D. Mesures des taux de refroidissement des chondrites ordinaires à partir des traces de fission du plutonium 244 enregistrées dans les cristaux détecteurs. *C. R. Acad. Sci. III D* **280**, 225–228 (1975).
- Trieloff, M., Deutsch, A. & Jessberger, E. K. The age of the Kara impact structure, Russia. *Meteoritics Planet. Sci.* **33**, 361–372 (1998).
- Schaeffer, G. A. & Schaeffer, O. A. ^{40}Ar – ^{39}Ar ages of lunar rocks. *Proc. Lunar Planet. Sci. Conf.* **8**, 2253–2300 (1977).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank B. Dominik for mineral separation, A. Bouikine and E. Korotchantseva for assisting in ^{40}Ar – ^{39}Ar analyses, and D. Stöffler for shock-stage classification of some H chondrites. We appreciate discussions with J.A. Wood, E. Anders and T. Althaus, and support from R. Altherr, T. Kirsten and K. Mauersberger at various stages of this study. M.T., E.K.J. and J.H. acknowledge support from the Deutsche Forschungsgemeinschaft.

Author contributions P.P. provided the motivation for this research by his enthusiastic work on ^{244}Pu fission-track studies.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.T. (e-mail: trieloff@min.uni-heidelberg.de).

Critical thickness for ferroelectricity in perovskite ultrathin films

Javier Junquera & Philippe Ghosez

Département de Physique, Université de Liège, Bâtiment B-5, B-4000 Sart-Tilman, Belgium

The integration of ferroelectric oxide films into microelectronic devices^{1,2}, combined with the size reduction constraints imposed by the semiconductor industry, have revived interest in the old question concerning the possible existence of a critical thickness for ferroelectricity. Current experimental techniques have allowed the detection of ferroelectricity in perovskite films down to a thickness of 40 Å (ten unit cells), ref. 3. Recent atomistic simulations^{4,5} have confirmed the possibility of retaining the ferroelectric ground state at ultralow thicknesses, and suggest the absence of a critical size. Here we report first-principles calculations on a realistic ferroelectric–electrode interface. We show that, contrary to current thought, BaTiO₃ thin films between two metallic SrRuO₃ electrodes in short circuit lose their ferroelectric properties below a critical thickness of about six unit cells (~24 Å). A depolarizing electrostatic field, caused by dipoles at the ferroelectric–metal interfaces, is the reason for the disappearance of the ferroelectric instability. Our results suggest the existence of a lower limit for the thickness of useful ferroelectric layers in electronic devices.

Epitaxial growth of high-quality thin films of ABO₃ perovskite oxides on semiconductors—Si (ref. 6) and Ge (ref. 1)—and metallic⁷ substrates has opened the door to the use of alternative materials in metal-oxide-semiconductor field-effect transistors (MOSFETs)^{1,8} and non-volatile ferroelectric random access memories (FERAMs)^{9,10}. Recently, 30 Gbit cm⁻² data storage densities have been demonstrated for Pb(Zr_{0.2}Ti_{0.8})O₃ films on a metallic oxide electrode². As the size of high-quality perovskite layers of interest for technological applications decreases, the fundamental questions of the thickness dependence of the ferroelectric properties, and of their possible disappearance at a finite critical thickness, become crucial.

It was thought that ferroelectricity was suppressed in small particles and thin films¹¹, but this situation changed after experiments that identified ferroelectric ground states in 40-Å-thick perovskite oxide films³ and in 10-Å-thick crystalline copolymer films¹². In support of these experimental observations, theoretical investigations using a phenomenological approach¹³, first-principles effective hamiltonian calculations⁴ and full first-principles simulations⁵ predicted ferroelectric ground states for various ABO₃ slabs under the condition of vanishing internal electric field. However, in all these approaches, the influence of a real metal–perovskite interface (including the finite screening length of the electrode, the interface chemistry and the strain conditions imposed by the substrate) on both the atomic relaxation and polarization was missing—or was only empirically included through the choice of specific electromechanical boundary conditions. The only more-fundamental study of metal–BaTiO₃ interfaces is due to Rao *et al.*¹⁴, but does not address the question of the ferroelectric instability.

Here we simulate from first principles the behaviour of a typical ferroelectric capacitor structure epitaxially grown on a thick SrTiO₃ substrate, and made of an ultrathin film of BaTiO₃ (a prototypical ferroelectric oxide) between two metallic SrRuO₃ electrodes¹⁵ (Fig. 1). The electrical boundary conditions are fixed by putting the electrodes in short-circuit. Calculations were done within density functional theory and the local density approximation.

We used a numerical atomic orbital method, as implemented in the SIESTA code¹⁶.

SrRuO₃ is an excellent electrode material for ferroelectric device structures¹⁷, because of the high crystalline quality and coherent interfaces obtained in heteroepitaxial growth with perovskite oxides^{18,19}. A SrO/TiO₂ interface with BaTiO₃ is expected experimentally (owing to the volatility of Ru) and assumed in our calculations. The simulations were carried out using a supercell technique. The basic unit, periodically repeated in space, is illustrated in Fig. 1 and corresponds to the generic formula [SrO-(RuO₂-SrO)_n/TiO₂-(BaO-TiO₂)_m], where *n* was fixed to five (a thickness large enough to avoid interaction between the two interfaces) and *m* ranged from two to ten. The periodic boundary conditions naturally impose the required short-circuit condition between the electrodes. The SrTiO₃ substrate was not explicitly treated, but was implicitly included by constraining the in-plane lattice constant of the whole structure to the bulk cell parameter of SrTiO₃ (ref. 20). Strain relaxations of bulk SrRuO₃ and BaTiO₃ were done under this constraint, and the resulting tetragonal unit cells were used as the building blocks for the supercell. At the bulk level, a compressive strain identical to that imposed by the substrate (~2%) is sufficient to suppress the in-plane ferroelectric instability of BaTiO₃, while the concomitant relaxation in the perpendicular direction amplifies²¹ the computed spontaneous polarization from 24 to 31 $\mu\text{C cm}^{-2}$.

The different heterostructures were first relaxed imposing a mirror symmetry plane on the central TiO₂ layer. The relaxation was continued until the maximum force on the atoms was smaller than 40 meV Å⁻¹. The symmetry constraint prevents the appearance of ferroelectricity, and the resulting structures correspond to the optimized paraelectric states. Starting from the previous relaxed configurations taken as reference, we then searched for the existence of a ferroelectric instability, moving BaTiO₃ atoms continuously following the displacement pattern of the bulk tetragonal soft mode²², ξ , determined for the same tetragonal cell geometry as in the supercell. Only the monodomain case was considered, corresponding to a situation that can be achieved experimentally⁹. In Fig. 2, the symbols show the evolution of the energy (with respect to the reference paraelectric phase) when freezing-in atomic displacements ξ of increasing amplitudes. Calculations have been done for different thicknesses of the BaTiO₃ layer. Figure 2 points out a change of behaviour at a critical thickness around *m* = 6 (26 Å). For

larger thicknesses, the energy decreases when the atoms follow the bulk soft-mode path, clearly demonstrating that the thin film has a ferroelectric ground state (in agreement with experimental results for the same thickness³). In contrast, below the critical value, the energy is minimized for the paraelectric configuration. Only a small region of the whole configuration space is analysed in Fig. 2, so this observation alone is not sufficient to make conclusions about the absence of ferroelectricity. However, as will become clear below, the mechanism responsible for the increase of energy when freezing-in the displacements depends only on the induced polarization and not on the details of the atomic displacement pattern. To further prove this point, we performed a full atomic relaxation for the *m* = 2 supercell, starting from a ferroelectric configuration, with the result that the atoms moved back to their paraelectric positions.

These first-principles results can be explained by simple electrostatic arguments. The polarization, **P**, generated in the ferroelectric thin film by a polar pattern of atomic displacements, leads to surface charges, $\sigma_{\text{pol}} = \mathbf{P} \cdot \hat{\mathbf{n}}$ (where $\hat{\mathbf{n}}$ is a unitary vector normal to the interface that points out of the ferroelectric film) of opposite sign at the interfaces with the electrodes. Electrons in the metal tend to screen these charges, and consequently two dipoles of the same sign arise at the two metal-ferroelectric interfaces of the sample. All these features are illustrated in Fig. 3a, where the difference in the macroscopic²³ total charge density distribution (ionic plus electronic) between the ferroelectric and the paraelectric phases is plotted. The induced deformation of the charge density in the metal has a complex structure, and decays exponentially in the metallic region beyond the first RuO₂ layer. It spreads over a characteristic distance that is related to the screening length of the electrode. Each dipole is responsible for electrostatic potential drops in the same direction, so that the only way to preserve short-circuit boundary conditions along the structure will be the appearance of a depolarizing electric field inside the ferroelectric film, illustrated in Fig. 3b for different values of the induced polarization. The amplitude of this field, $\mathcal{E}_d$, depends on (1) the screening length of the metal and the polarization of the thin film, which both monitor

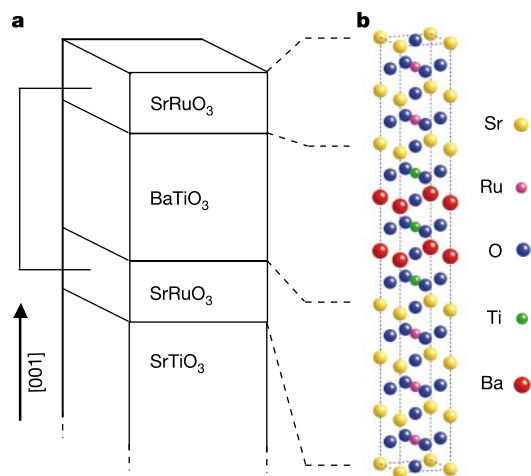


Figure 1 Structure of a typical ferroelectric capacitor. **a**, Schematic view of the system considered here: a BaTiO₃ film between SrRuO₃ electrodes that are short-circuited. The whole structure is grown on top of a SrTiO₃ substrate. The epitaxy is such that SrTiO₃ (001) || SrRuO₃ (001) || BaTiO₃ (001) || SrRuO₃ (001). **b**, Atomic representation of the related supercell that is simulated (for the case *m* = 2).

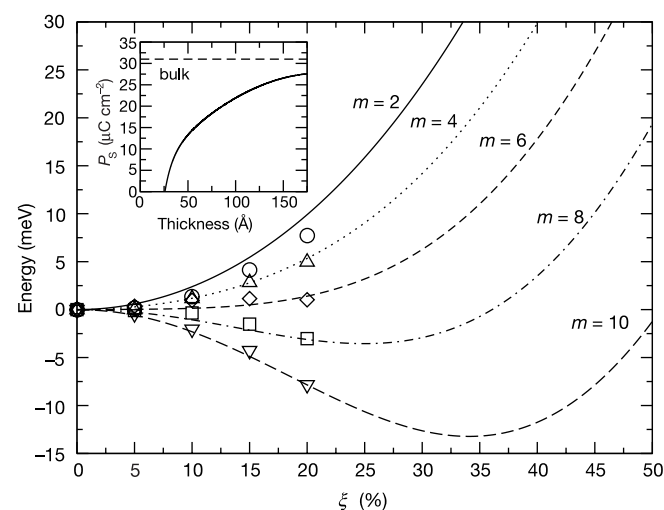


Figure 2 Evolution of the energy as a function of the soft-mode distortion ξ . First-principles results (symbols) and electrostatic model results (lines) are shown for different thicknesses of the ferroelectric thin film: *m* = 2 (circles, full line), *m* = 4 (upward-pointing triangles, dotted), *m* = 6 (diamonds, short-dashed), *m* = 8 (squares, dot-dashed), and *m* = 10 (downward-pointing triangles, long-dashed). The magnitude of ξ is reported as a percentage of the bulk soft-mode displacements of BaTiO₃ atoms ($\xi = 1$ corresponds to the distortion of the bulk tetragonal ferroelectric phase). The energy of the paraelectric phase is taken as reference. Inset, evolution of the spontaneous polarization, P_s , with thickness. For each thickness, P_s is deduced from the amplitude of ξ at the minimum.

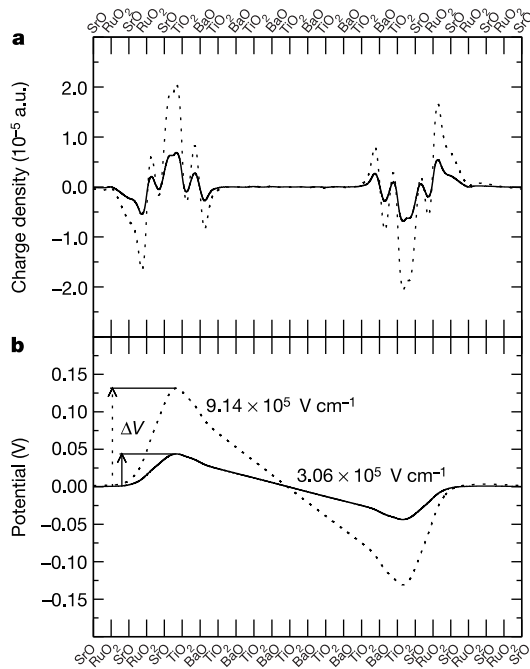


Figure 3 Induced dipoles, potentials and depolarizing fields along the [001] direction of the short-circuited ferroelectric capacitor. Shown are the change in **a**, the macroscopic total (ionic plus electronic) charge density (in atomic units, a.u.), and **b**, the electrostatic potential with respect to the paraelectric phase, when freezing-in the soft-mode distortion pattern: $\xi = 5\%$ (full line) and $\xi = 15\%$ (dotted line). The thickness of the BaTiO₃ layer corresponds to $m = 6$. The magnitude of the depolarizing field for each configuration is indicated. ΔV , the potential drop across the interface.

the potential drop ΔV across the interface, and (2) the film thickness l ($\mathcal{E}_d = 2\Delta V/l$). The effect of the depolarizing field on the energy of the system can be extrapolated from the following model.

In the presence of the field $\mathcal{E}_d$, the energy E of the thin film can be expressed as²⁴ $E = U - \mathbf{P} \cdot \mathcal{E}_d$, where U is the internal energy under zero field. Assuming that the modification of the bonding at the interface only plays a minor role, the evolution of U can be approximated from the bulk soft-mode double-well energy²². To calculate $\mathbf{P}$, contributions from the atomic displacements (ξ_i^α , where i is an atomic index and α is a cartesian direction) and from the electronic polarization in the presence of the field are taken into account ($P_\beta = (1/\Omega) \sum_{i,\alpha} Z_{\alpha\beta,i}^{(*)} \xi_i^\alpha + \chi^\infty \mathcal{E}_{d\beta}$, where Ω is the volume of the ferroelectric thin film). Again, neglecting deviations in the surface region, bulk values can be assumed for the different quantities. The transverse effective charges $Z_{\alpha\beta,i}^{(*)}$ were calculated with the Berry phase approach²⁵, whereas the electronic susceptibility χ^∞ was obtained from a supercell technique²⁶. Finally, the amplitude of the field $\mathcal{E}_d$ can be deduced from the slope of the macroscopic average of the electrostatic potential in the ferroelectric (Fig. 3b).

With this simplified model (lines in Fig. 2), all the first-principles results are reproduced. This shows that the behaviour of thin ferroelectric films is dominated by electrostatics whereas, for the structure under investigation, the chemistry of the interface (neglected in the model) only plays a minor role. In the case of zero internal field (assuming perfect screening⁴ or compensating for the depolarizing field⁵), E reduces to U and the behaviour of the film is similar to that of the bulk. However, in the case of a real metal–ferroelectric interface, perfect screening is not achieved under short-circuit boundary conditions, resulting in a sizeable depolarizing field. The existence of such a field was previously suggested by Mehta *et al.*²⁷, and was recently invoked to explain experimental changes of the coercive fields with thickness²⁸. Here, we quantify the

effect of the depolarizing field for a realistic system and demonstrate its crucial role. First, it is responsible for the suppression of ferroelectricity at a finite size. Second, even well above the critical thickness, the depolarizing field reduces the amplitude of the spontaneous polarization, which only slowly approaches the bulk value for films of increasing size (Fig. 2 inset).

The results reported here have been obtained for a defect-free perfectly insulating BaTiO₃ crystal, whereas typical ferroelectric capacitors are well-known to exhibit high leakage currents (mainly monitored by the Poole–Frenkel mechanism in the case of metal oxide electrodes²⁹). It might therefore be questioned whether charge redistribution associated with the finite conductivity of the ferroelectric could further screen the depolarizing field and affect our conclusions. For usual current densities³⁰ ($J \approx 10^{-5} \text{ A cm}^{-2}$, at 300 K and for fields of the order of those of Fig. 3b), screening of the uncompensated surface polarization charges ($\sigma_{\text{pol}} \approx 10^{-6} \text{ C cm}^{-2}$) would require a timescale of the order of 0.1 s. Compared to the switching of the polarization (typically $\sim 10^{-9}$ s), this is a very slow process. We can therefore conclude that this process cannot prevent the film switching back to the paraelectric state below the critical thickness. For m larger than the critical value and after a sufficiently long time, the finite conductivity could allow the appearance of additional screening charges. However, as it happens for charges coming from the electrodes, the finite spread of this extra charge distribution will probably provide an incomplete screening. A non-vanishing depolarizing field is therefore still expected, and might have marked effects on fundamental properties important for technological applications, such as the coercive fields²⁸ (and hence the switching voltage) or fatigue of the samples.

We believe that our results will stimulate further work on growing ultrathin perovskite films, and on measuring the thickness dependence of the polarization within such films. □

Received 9 September 2002; accepted 12 February 2003; doi:10.1038/nature01501.

- McKee, R. A., Walker, F. J. & Chisholm, M. F. Physical structure and inversion charge at a semiconductor with a crystalline oxide. *Science* **293**, 468–471 (2001).
- Tybell, Th., Paruch, P., Giamarchi, T. & Triscone, J.-M. Domain wall creep in epitaxial ferroelectric Pb(Zr_{0.2}Ti_{0.8})O₃ thin films. *Phys. Rev. Lett.* **89**, 097601 (2002).
- Tybell, Th., Ahn, C. H. & Triscone, J.-M. Ferroelectricity in thin perovskite films. *Appl. Phys. Lett.* **75**, 856–858 (1999).
- Ghosez, Ph. & Rabe, K. M. Microscopic model of ferroelectricity in stress-free PbTiO₃ ultrathin films. *Appl. Phys. Lett.* **76**, 2767–2769 (2000).
- Meyer, B. & Vanderbilt, D. Ab initio study of BaTiO₃ and PbTiO₃ surfaces in external electric fields. *Phys. Rev. B* **63**, 205426 (2001).
- McKee, R. A., Walker, F. J. & Chisholm, M. F. Crystalline oxides on silicon: the first five monolayers. *Phys. Rev. Lett.* **81**, 3014–3017 (1998).
- Ahn, C. H. *et al.* Local, nonvolatile electronic writing of epitaxial Pb(Zr_{0.52}Ti_{0.48})O₃/SrRuO₃ heterostructures. *Science* **276**, 1100–1103 (1997).
- Eisenbeiser, K. *et al.* Field effect transistors with SrTiO₃ gate dielectric on Si. *Appl. Phys. Lett.* **76**, 1324–1326 (2000).
- Tybell, Th., Ahn, C. H. & Triscone, J.-M. Control and imaging of ferroelectric domains over large areas with nanometer resolution in atomically smooth epitaxial Pb(Zr_{0.52}Ti_{0.48})O₃ thin films. *Appl. Phys. Lett.* **72**, 1454–1456 (1998).
- Scott, J. F. & Paz de Araujo, C. A. Ferroelectric memories. *Science* **246**, 1400–1405 (1989).
- Llines, M. E. & Glass, A. M. *Principles and Applications of Ferroelectrics and Related Materials* (Clarendon, Oxford, 1977).
- Bune, A. V. *et al.* Two-dimensional ferroelectric films. *Nature* **391**, 874–877 (1998).
- Kohlstedt, H., Pertsev, N. A. & Waser, R. Size effects on polarization in epitaxial ferroelectric films and the concept of ferroelectric tunnel junctions including first results. *Mater. Res. Soc. Symp. Proc.* **688**, C6.5 (2002).
- Rao, E., Kim, M., Freeman, A. J., Tang, S. & Anthony, M. Structural and electronic properties of transition-metal/BaTiO₃ (001) interfaces. *Phys. Rev. B* **55**, 13953–13960 (1997).
- Eom, C. B. *et al.* Single-crystal epitaxial thin films of the isotopic metallic oxides Sr_{1-x}Ca_xRuO₃ ($0 \leq x \leq 1$). *Science* **258**, 1766–1769 (1992).
- Soler, J. M. *et al.* The SIESTA method for ab initio order-N materials simulations. *J. Phys. Condens. Matter* **14**, 2745–2779 (2002).
- Hartmann, A. J., Neilson, M., Lamb, R. N., Watanabe, K. & Scott, J. F. Ruthenium oxide and strontium ruthenate electrodes for ferroelectric thin-film capacitors. *Appl. Phys. A* **70**, 239–242 (2000).
- Ahn, C. H. *et al.* Ferroelectric field effect in ultrathin SrRuO₃ films. *Appl. Phys. Lett.* **70**, 206–208 (1997).
- Eom, C. B. *et al.* Fabrication and properties of epitaxial ferroelectric heterostructures with (SrRuO₃) isotropic metallic oxide electrodes. *Appl. Phys. Lett.* **63**, 2570–2572 (1993).
- Zimmer, M., Junquera, J. & Ghosez, Ph. Ab initio study of metal/ferroelectric and insulator/ferroelectric heterostructures. *AIP Conf. Proc.* **626**, 232–241 (2002).
- Neaton, J. B., Hsueh, C. L. & Rabe, K. M. Enhanced polarization in strained BaTiO₃ from first principles. *Mater. Res. Soc. Symp. Proc.* **718**, 311 (2002).

22. Cohen, R. E. Origin of ferroelectricity in perovskite oxides. *Nature* **358**, 136–138 (1992).
23. Baldereschi, A., Baroni, S. & Resta, R. Band offsets in lattice-matched heterojunctions: a model and first-principles calculations for GaAs/AlAs. *Phys. Rev. Lett.* **61**, 734–737 (1988).
24. Fu, H. & Cohen, R. E. Polarization rotation mechanism for ultrahigh electromechanical response in single-crystal piezoelectrics. *Nature* **403**, 281–283 (2000).
25. King-Smith, R. D. & Vanderbilt, D. Theory of polarization of crystalline solids. *Phys. Rev. B* **47**, 1651–1654 (1993).
26. Bernardini, F. & Fiorentini, V. Electronic dielectric constants of insulators calculated by the polarization method. *Phys. Rev. B* **58**, 15292–15295 (1998).
27. Mehta, R. R., Silverman, B. D. & Jacobs, J. T. Depolarization fields in thin ferroelectric films. *J. Appl. Phys.* **44**, 3379–3385 (1973).
28. Chandra, P., Dawber, M., Littlewood, P. B. & Scott, J. F. Thickness-dependence of the coercive field in ferroelectrics. Preprint at (<http://xxx.lanl.gov/abs/cond-mat/0206014>) (2002).
29. Nagaraj, B., Aggarwal, S. & Ramesh, R. Influence of contact electrodes on leakage characteristics in ferroelectric thin films. *J. Appl. Phys.* **90**, 375–382 (2001).
30. Hwang, C. S. *et al.* A comparative study on the electrical conduction mechanisms of (Ba_{0.5}Sr_{0.5})/TiO₃ thin films on Pt and IrO₂ electrodes. *J. Appl. Phys.* **83**, 3703–3713 (1998).

Acknowledgements We thank Th. Tybell, J.-M. Triscone, H. Kohlstedt, M. Veithen, X. Gonze and D. Vanderbilt for discussions, and K. M. Rabe and A. Garcia for critical reading of the manuscript. This work was supported by the VolkswagenStiftung, the ULG, the FNRS-Belgium, the CGRI, the Fundación Ramón Areces and the Spanish MCYT.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.G. (e-mail: Philippe.Ghosez@ulg.ac.be).

Homogeneous climate variability across East Antarctica over the past three glacial cycles

O. Watanabe*, J. Jouzel†, S. Johnsen‡, F. Parrenin§, H. Shoji|| & N. Yoshida¶#

* National Institute of Polar Research, 9-10 Kaga 1-chome, Itabashi-ku, Tokyo 173, Japan

† IPSL/Laboratoire des Sciences du Climat et de l'Environnement, UMR CEA-CNRS, CE Saclay, 91191 Gif-sur-Yvette, France

‡ Department of Geophysics, Juliane Maries Vej 30, University of Copenhagen, DK-2100, Copenhagen, Denmark, and Science Institute, University of Reykjavik, Dunhaga 3, Reykjavik 107, Iceland

§ Laboratoire de Glaciologie et Géophysique de l'Environnement, CNRS, BP 96, 38402, Saint Martin d'Hères, Cedex, France

|| New Energy Resources Research Center, Kitami Institute of Technology, 165, Koen-Cho, Kitami 090-8507, Japan

¶ Frontier Collaborative Research Center, Tokyo Institute of Technology, 4259, Nagatsuta, Midori-ku, Yokohama 226-8502, Japan

The SORST Project, Japan, Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan

Recent ice core studies have raised the disturbing possibility that glacial–interglacial climate changes may be non-uniform across Antarctica^{1,2}. These findings have been confined to records from the Ross Sea sector of the continent, but significant deviations in other areas would call into question the widely assumed validity of the climate record obtained from Vostok, East Antarctica, on large spatial scales³. Here we present an isotopic profile from a core drilled at Dome Fuji^{4,5}, situated 1,500 km from Vostok in a different sector of East Antarctica. The two records show remarkable similarities over the past three glacial cycles (the extent of the Dome Fuji record) in both large-amplitude changes, such as terminations, interglacials and interstadials and more subtle glacial events, even when the origin of precipitation is accounted for. Our results indicate that Antarctic climate is essentially homogeneous at the scale of the East Antarctic Plateau, possibly

as a consequence of the symmetry of the plateau and the adjacent ocean.

Climate records extending at least back to the Last Glacial Maximum, 20 kyr BP (thousands of years before present), have now been recovered from deep ice cores drilled at seven different Antarctic sites (Fig. 1). The inland East Antarctic records (Dome C, Dome B and Vostok, with two published series at Dome C and Vostok) are very similar where they overlap, with a deglaciation history characterized by a long gradual warming interrupted by the Antarctic Cold Reversal, clearly preceding the North Atlantic Younger Dryas⁶. These characteristics are shared by the records from Byrd⁷ (inland West Antarctica) and Law Dome (coastal East Antarctica)⁸. Although the exact timing of the Antarctic Cold Reversal is still debated⁸, all these records show a deglacial pattern very different from that registered in Greenland cores. In contrast, the records from Taylor Dome¹ and Siple Dome², two near-coastal sites in the Ross Sea sector (Fig. 1), show Greenland-like features with indication of either climate changes synchronous with the North Atlantic¹ or of very rapid changes typical of the Greenland record².

Here we present a climate isotopic record from a core successfully recovered by Japanese drillers in 1995 and 1996 at Dome Fuji in East Antarctica⁴, which, in this context, is important for at least two reasons. First, Dome Fuji and other inland East Antarctic sites are located in completely different sectors of East Antarctica (Fig. 1). Second, the Dome Fuji core extends back to about 330 kyr BP (ref. 9). It provides the second-longest climate time series after the Vostok cores, allowing a detailed comparison over the last three glacial–interglacial cycles, including three successive deglaciations (Fig. 2). Ongoing Dome Fuji studies provide a wealth of new and detailed information in various fields (climate, atmospheric chemistry, atmospheric composition, ice properties and ice dynamics).

We focus on the continuous ¹⁸O profile in ice (δ¹⁸O) and on its comparison with the Vostok deuterium record³ (δD). We also shortly discuss implications of the two deuterium-excess profiles ($d = \delta D - 8 \times \delta^{18}O$) in terms of local temperature interpretation.

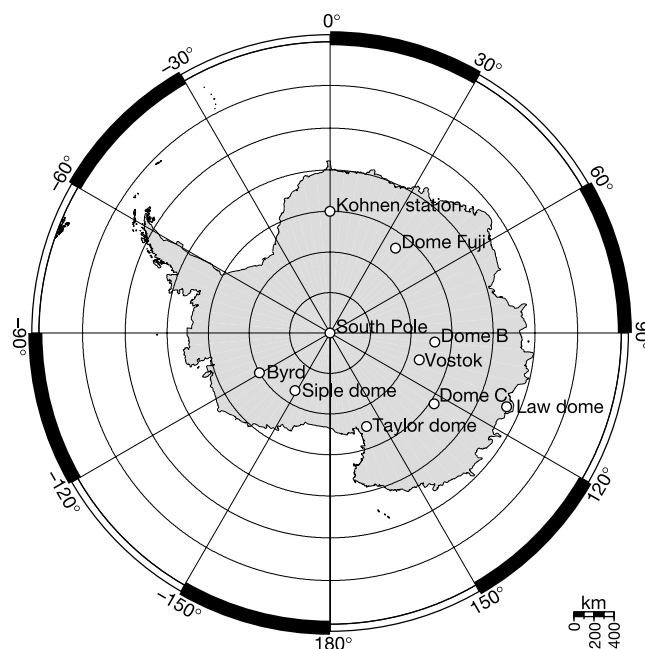


Figure 1 Map of Antarctica with indication of the deep drilling sites. The Dome Fuji site is located about 1,000 km from the coast at the highest point of the East Dronning Maud Land Plateau: 3,810 m above sea level (a.s.l.), accumulation $\sim 2.7 \text{ g cm}^{-2} \text{ yr}^{-1}$, 10-m depth temperature of -58°C (ref. 9). The core covers 2,503 m, whereas the ice thickness is 3,090 m.

The similarities between the inferred Dome Fuji and Vostok climate records, which hold for the whole period and even for very detailed features, are remarkable given the distance between the two sites (about 1,500 km) and a plausible different precipitation origin.

Glaciologists rely on the δD or $\delta^{18}O$ composition of the ice to reconstruct continuous records of past temperature changes, ΔT_s . The two parameters are linearly related with a slope close to eight¹⁰ and both may be used interchangeably. A different choice was made by the Vostok and Dome Fuji teams who used δD and $\delta^{18}O$ for temperature reconstruction (see Methods), respectively. Figure 2 shows these isotopic profiles as a function of depth and age, whereas inferred ΔT_s values are compared in Fig. 3. The deepest Dome Fuji samples correspond to the last part of termination IV, dated at about 340 kyr BP in the deep-sea core chronology¹¹.

To compare the Dome Fuji and Vostok climatic records, we placed them on a common timescale. We adopted the Dome Fuji timescale obtained by the inverse method developed in ref. 12, an approach ideally suited for dating a core located on a dome (see Methods). To transfer the Vostok record onto the Dome Fuji timescale, numerous tie points can be defined but we prefer to use only a few and interpolate between them (Fig. 2), to examine how other events correlate. The similarity between the two records is remarkable. Large interglacials, interstadials and smaller events are quite similar both in shape and amplitude in the two records. There are a few noticeable exceptions. The low values around 45 kyr BP are more pronounced at Vostok than at Dome Fuji and vice versa around 100 kyr BP. The peak just before 175 kyr BP is clearly marked at Dome Fuji and muted at Vostok whereas the situation is reversed

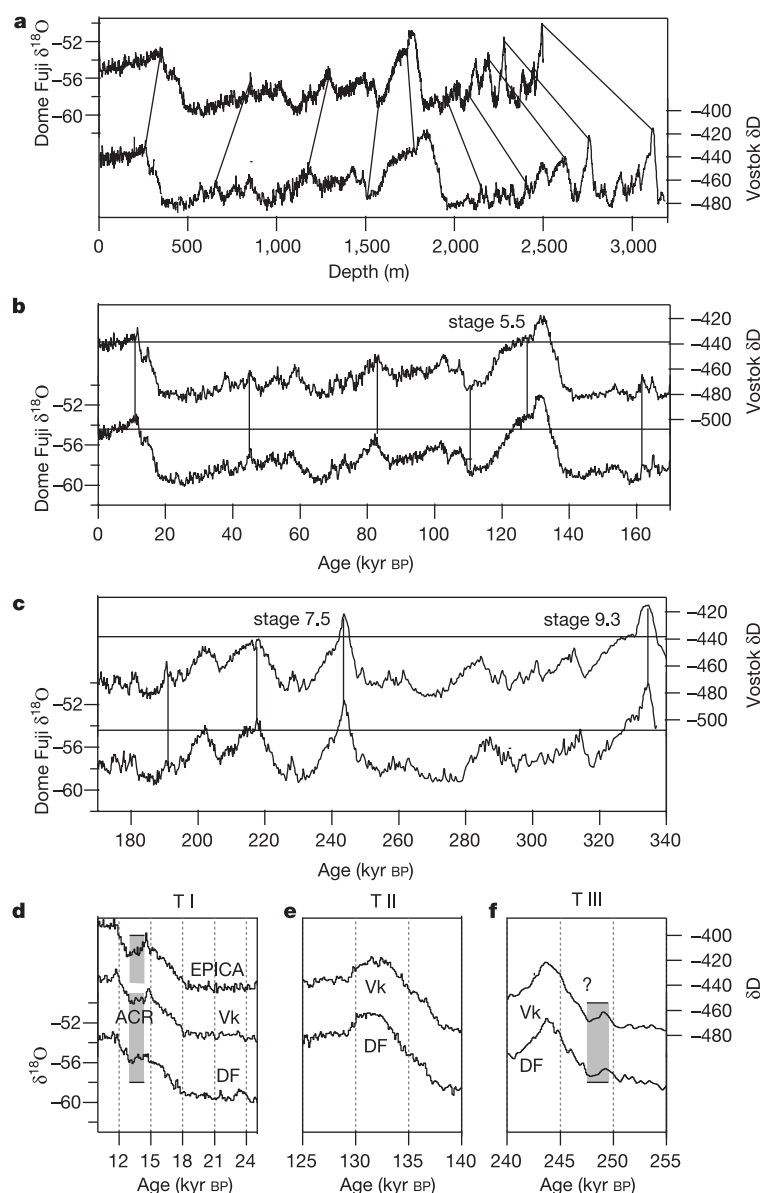


Figure 2 Comparison of the Vostok and Dome Fuji isotopic records as a function of depth and time. For the sake of visual comparison δD and $\delta^{18}O$, expressed in per mil units with respect to VSMOW, the Vienna Standard Mean Ocean Water) are scaled in a ratio of 7.94 which corresponds to the Vostok $\delta D/\delta^{18}O$ slope¹⁰. We use the inverse Dome Fuji timescale on which the Vostok record is transferred using the tie points of **a**, and a linear interpolation with respect to age rather than to depth because the thinning function differs between the two sites. For this comparison the two records have been resampled at

equally spaced intervals of 100 years. Vertical lines (in **b** and **c**) correspond to the tie points shown in **a**, and horizontal lines correspond to the mean Holocene isotopic content for each core. The Vostok deuterium profile is shown down to 3,200 m in order to include the end of marine stage 10 and termination IV. **d–f**, Expanded views of terminations I, II and III. For termination I (**d**), we have also reported the EPICA Dome C deuterium record⁶. ACR, Antarctic Cold Reversal; Vk, Vostok; DF, Dome Fuji.

for the peak just before 300 kyr BP. The coldest part of the third climatic cycle seems to start about 7 kyr earlier at Dome Fuji than at Vostok but this is clearly due to the small number of tie points. We could list other examples but overall those differences are relatively minor and might even be readily explained by depositional noise, in such low accumulation areas.

At Vostok the coldest part of each glacial period occurred just before the termination with the noticeable exception of the third cycle. Termination III, starting at ~248 kyr BP for its main rise, is preceded by a series of three short events. Exactly the same sequence is observed at Dome Fuji with a small but well-marked event between 251 and 248 kyr BP. As a result, we can consider termination III (Fig. 2f) to be a two-step transition similar to termination I (Fig. 2d) but with a Cold Reversal at the very beginning of this transition; we would expect confirmation either from other properties or from another core such as EPICA Dome C¹³. In contrast, termination II (Fig. 2e) is very regular both at Vostok and Dome Fuji where termination I (Fig. 2d) shows a two-step warming with a first trend ending slightly before 14.5 kyr BP. The Antarctic Cold Reversal has its coldest part just before 13 kyr BP and is then followed by a second warming trend starting around 13 kyr BP, about 1.5 kyr before the end of the Younger Dryas. Similarities with other inland East Antarctic cores extend to the early Holocene epoch, characterized by a clear optimum at 12–9 kyr BP (ref. 14).

At both sites, warm interglacials (marine stages 9.3, 7.5 and 5.5) are followed by increasingly colder interstadial events that end with a rapid return towards the following interglacial. As in the Vostok record, the Dome Fuji record confirms that interglacials 9.3 and 5.5 are very similar in duration, shape and amplitude but differ both from the Holocene and from interglacial 7.5. This is illustrated in Fig. 3a, which compares the two isotopic records directly in terms of local temperature, applying the conventional approach based on present-day observations^{3,15} (see Methods). Unlike for the Greenland record, the use of this spatial slope as a surrogate of the temporal slope in central East Antarctica appears justified within $\pm 20\%$ (ref. 15).

As reviewed in ref. 15, this is largely because glacial–interglacial changes in the seasonality and in the origin of the precipitation have no strong influence on the temporal slope. Both aspects have been

examined using a variety of modelling approaches^{16,17}, and this result is supported overall by experiments performed with isotopic General Circulation Models^{15,18}. Even when such experiments show noticeable differences in the oceanic origin of the Vostok and Dome Fuji precipitation, they do not seem greatly to affect ΔT_s estimates¹⁷. This can be independently evaluated when both δD and $\delta^{18}O$ profiles are available, with a correction expressed as a function of the change in deuterium-excess^{10,19}. Because of less continuous sampling at Dome Fuji at present, the excess record shows much higher variability than at Vostok and we thus compare, in Fig. 3, smooth versions of this parameter. These profiles show glacial–interglacial changes of relatively similar amplitude and shape. In turn, accounting for this source correction, which is in any case relatively small, does not significantly alter the similarities between estimated local temperature changes (see Methods).

As for the Vostok record, the peaks of stages 9.3, 7.5 and 5.5 were warmer at Dome Fuji than during the Holocene, but they appear also to be systematically warmer than the corresponding Vostok peaks (by more than 1 °C for stage 9.3). Dome Fuji should provide a better record than Vostok where deeper ice originates more than 200 km upstream of the drilling site. Although an explanation based on an elevation change cannot be definitely ruled out, we tentatively assumed that this Dome Fuji / Vostok surface temperature difference relates to the origin of the Vostok ice, and applied a slight correction to the Vostok profile accordingly (see Methods). In turn, interglacial 9.3 was probably the warmest period over the last 400 kyr in Antarctica with maximum surface temperature up to 6 °C warmer than during the Holocene. Not only is there a high degree of similarity between the two temperature records (Fig. 3a) given the distance between the two sites (with a correlation coefficient R^2 of 0.96), but also the overall amplitude of change is exactly the same at both sites.

This similarity further reinforces the validity of using the conventional approach because if its application were biased (either at Vostok or at Dome Fuji, or at both sites) owing to changes in precipitation seasonality, such variation would not be completely uniform over a spatial scale of 1,500 km. Alternatively, these strong isotopic similarities may occur by chance; we cannot definitely exclude the possibility that ΔT_s does differ between the two sites but that the isotopic records nevertheless have a similar amplitude owing to some compensation effect. Whatever the correct explanation, Fig. 3a clearly supports the conclusion³ that the broad features of the Vostok record are of geographical significance for a large area (Antarctica and part of the Southern Hemisphere).

The present study thus considerably extends previous work both geographically and temporally. Dome Fuji and Vostok records show a remarkably high degree of similarity given that three climatic cycles are depicted at two sites from different sectors of East Antarctica. Successive terminations and large interglacial events as well as intervening large interstadials have comparable shape. This applies to termination I, strongly suggesting that Greenland-like features^{1,2} are, in any case, restricted to the Ross Sea sector. Indeed, the Dome Fuji record adds considerably to the geographical significance of the Antarctic Cold Reversal event, until now based mainly on cores located in a sector representing about one-third of East Antarctica⁶. This also applies to smaller events which, for the last glacial period^{20,21}, are counterparts of the Dansgaard–Oeschger events recorded in the Greenland records²². Dome Fuji is closer to the Atlantic Ocean than Vostok, so it is somewhat surprising that those Dansgaard–Oeschger events are not recorded differently in the two sites. However, model experiments in which the origin of water vapour is tagged^{23,24} suggest that, at both sites, a significant part of the precipitation originates from the Indian and Pacific oceans. They thus provide a reasonable explanation for observed similarities and reinforce the need for further drillings at sites in the Atlantic sector.

As noted, the comparison between Dome Fuji and Vostok

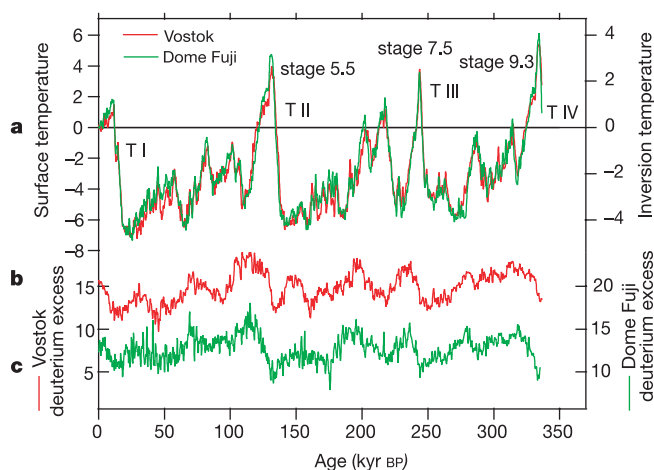


Figure 3 Comparison of the Dome Fuji and Vostok isotopic temperature and deuterium-excess^{10,25} records. **a**, The surface, ΔT_s , and inversion, ΔT_i , temperature (°C) differences, with $\Delta T_i = 0.67\Delta T_s$ (ref. 15) with respect to the last 1,000 years. The curves are smoothed versions of the temperature records inferred from the isotopic records (see Methods) with further small time adjustments with respect to Fig. 2b, c. Terminations I to IV and marine stages are shown. **b**, **c**, The smoothed Vostok and Dome F deuterium-excess records (‰ VSMOW). All these curves are obtained using an 11-points least-squares smoothing performed on a 100-yr resampled time series.

temperature records indirectly supports the conventional interpretation of isotopic records for Antarctica. It suggests that stages 9.5 and 5.5 were much warmer than the most recent 1,000 years ($\sim 4.5^\circ\text{C}$ for stage 5.5 and up to 6°C for stage 9.3). More importantly, it suggests overall that the Antarctic climate is essentially homogeneous when examined at large geographical (East Antarctic Plateau) and temporal scales. We suggest that this homogeneity of the Antarctic climate at these scales might result from the relative symmetry of East Antarctica and of the adjacent Southern Ocean. However, we expect that examining other ice-core properties will reveal differences due to the oceanic origin of the air masses that provide precipitation at each site.

It will also be interesting to extend this comparison to other Antarctic cores, in particular the Dome C and Dronning Maud Land cores (Fig. 1) currently drilled in the framework of EPICA (European Project for Ice Drilling in Antarctica). Together with Dome Fuji, the Dome C core, which may already cover 520 kyr (ref. 13), will strongly constrain the duration of successive events (such as the last interglacial) and in turn the Antarctic chronology, whereas the Dronning Maud Land core is ideally located (about 1,000 km from Dome Fuji and closer to the Atlantic) to make the link with the Greenland records. $\square$

Methods

The $\delta^{18}\text{O}$ concentration was measured over the full length of the Dome Fuji core at the National Institute for Polar Research⁹ with a depth resolution increasing from its top to its bottom (ice increments from $\sim 2\text{ m}$ for the Holocene period to 10 cm for the third climatic cycle). As a result, the time resolution is, for periods older than $\sim 100\text{ kyr BP}$, better or much better at Dome Fuji than at Vostok, where δD measurements were performed every metre. For example, the third climatic cycle between interglacials 9.3 and 7.5 is represented by 400 samples (average resolution of 200 years) at Vostok and by $\sim 2,000$ samples at Dome Fuji (average resolution of 40 years). The deuterium-excess Dome F profile was obtained from a partly different set of less continuous samples ($10\text{--}25\text{ cm}$ every 1.85 m on average), measured at the Tokyo Institute of Technology²⁵.

The inverse dating method¹² combines an accumulation history and an ice flow model, under the assumption that the number of precessional cycles can be correctly counted. This assumption is straightforward for the Vostok deuterium record as well as for the record of atmospheric ^{18}O , strongly influenced by the precession³. Rather than having fixed control points³, the inverse dating method accounts for the uncertainty associated with any chronological information and then identifies glaciological chronologies which best fit this information. It was developed for dating the Vostok ice core, where, however, it fails to provide a reliable timescale beyond the last two climatic cycles. This is probably because accumulation changes upstream of Vostok, from where the Vostok ice originates, are poorly known¹².

In contrast, the Dome Fuji core is ideally suited to develop such a timescale as, unlike for the Vostok record, ice found at depth is formed from snow that accumulates at the site itself. The bottom of the core is also about 600 m above the bedrock, which ensures that the records obtained are undisturbed. However, we can at present rely on orbital constraints derived from the ice $\delta^{18}\text{O}$ record only. The inverse Dome Fuji timescale should thus be considered preliminary, as it would be useful to account also for the information contained in the air-bubble ^{18}O . However, we use the timescale here because it is of equal quality over the entire record, confirming for example that the GT4 Vostok chronology is too young for the third climatic cycle³. The isotope and temperature records are presented using this Dome Fuji inverse timescale (Figs 2 and 3).

The choice of the timescale is indeed not central for the comparison between Dome Fuji and Vostok climate records on which we focus here. We note that the duration of the last interglacial (stage 5.5) is shorter than in previous Vostok chronologies^{3,26}, that is, about 14 kyr at mid-transition (Fig. 2d). As for the inverse Vostok dating, the age at mid-transition ($\sim 134\text{ kyr BP}$) is significantly younger than that of 139 kyr BP derived in the initial Vostok dating²⁶ and thus does not support recent claims for an early termination I^{27,28}.

The surface-temperature changes displayed in Fig. 3a are derived as follows. Present-day observations show $\delta^{18}\text{O}/T_s$ spatial slopes of $0.852\text{‰ per }^\circ\text{C}$ in the Dome Fuji sector and of $0.76\text{‰ per }^\circ\text{C}$ ($0.64/0.794$)^{10,15} in the Vostok sector, where a recent survey, however, provides an estimate $\sim 15\%$ higher²⁹. We thus considered that $\delta^{18}\text{O}/T_s$ slopes are similar in each sector within estimated uncertainties ($\pm 10\%$). After correcting for the ocean isotopic change³⁰, $\Delta\delta^{18}\text{O}_{\text{ocean}}$, we used a common value of $0.8\text{‰ per }^\circ\text{C}$ to infer ΔT_s . A small corrective term is then added to the Vostok ΔT_s in order to account for the upstream origin of the Vostok ice; this correction (maximum value of 1.2°C at 330 kyr BP) is taken proportional to the distance upstream from Vostok. We then examined the possible influence of changes in oceanic sources combining information from both isotopes and inverting a simple isotopic model^{10,19}. For the Vostok site¹⁰, this method leads to $\Delta T_s = 0.176 \times \Delta\delta\text{D} + 0.501 \times \Delta d + 0.558 \times \Delta\delta^{18}\text{O}_{\text{ocean}}$ and we provisionally used the same formula for Dome Fuji. The comparison between the temperature records is essentially unaltered ($R^2 = 0.95$ instead of 0.96 for the curves displayed in Fig. 3a).

As a sensitivity test, we then applied the inversion formula derived for EPICA Dome C¹⁹, $\Delta T_s = 0.161 \times \Delta\delta\text{D} + 0.446 \times \Delta d + 0.354 \times \Delta\delta^{18}\text{O}_{\text{ocean}}$, to both the Dome Fuji and

Vostok records. This alternative approach leads to a similar conclusion ($R^2 = 0.96$). This temperature interpretation should nevertheless be considered preliminary until the inversion procedure has been applied to the Dome Fuji records. This might lead to a different $\delta^{18}\text{O}/T_s$ slope and reveal differences in the source correction between Vostok and Dome Fuji. However, although this would affect the ΔT_s amplitude it will not have a significant effect on the Dome Fuji/Vostok correlation.

Received 15 November 2002; accepted 20 February 2003; doi:10.1038/nature01525.

- Steig, E. *et al.* Synchronous climate changes in Antarctica and the North Atlantic. *Science* **282**, 92–95 (1998).
- Severinghaus, J. P., Grachev, A., Luz, B. & Caillon, N. A method for precise measurement of argon $40/36$ and krypton/argon ratios in trapped air in polar ice with applications to past firm thickness and abrupt climate change in Greenland and at Siple Dome, Antarctica. *Geochim. Cosmochim. Acta* **67**, 325–343 (2003).
- Petit, J. R. *et al.* Climate and atmospheric history of the past 42000 years from the Vostok Ice Core, Antarctica. *Nature* **399**, 429–436 (1999).
- Dome F Ice Core Research Group. Deep-ice core drilling at Dome Fuji and glaciological studies in east Dronning Maud Land, Antarctica. *Ann. Glaciol.* **27**, 333–337 (1998).
- Dome F Ice Core Research Group. Preliminary investigation of paleoclimate signals recorded in the ice core from Dome Fuji station, East Dronning Maud Land, Antarctica. *Ann. Glaciol.* **27**, 338–342 (1998).
- Jouzel, J. *et al.* A new 27 kyr high resolution East Antarctic climate record. *Geophys. Res. Lett.* **28**, 3199–3203 (2001).
- Johnsen, S. J., Dansgaard, W., Clausen, H. B. & Langway, C. C. J. Oxygen isotope profiles through the Antarctic and Greenland ice sheets. *Nature* **235**, 429–434 (1972).
- Morgan, V. *et al.* Relative timing of deglacial climate events in Antarctica and Greenland. *Science* **297**, 1862–1864 (2002).
- Watanabe, O. *et al.* The paleoclimate record in the ice core at Dome Fuji station, East Antarctica. *Ann. Glaciol.* **29**, 176–178 (1999).
- Vimeux, F., Cuffey, K. & Jouzel, J. New insights into Southern Hemisphere temperature changes from Vostok ice cores using deuterium excess correction over the last 420,000 years. *Earth Planet. Sci. Lett.* **203**, 829–843 (2002).
- Bassinot, F. C. *et al.* The astronomical theory of climate and the age of the brunhes-matuyama magnetic reversal. *Earth Planet. Sci. Lett.* **126**, 91–108 (1994).
- Parrenin, F., Jouzel, J., Waelbroeck, C., Ritz, C. & Barnola, J. M. Dating the Vostok ice core by an inverse method. *J. Geophys. Res.* **106**, 31837–31851 (2001).
- EPICA Dome C 2001–02 Science And Drilling Teams. Extending the ice core record beyond half a million years. *Eos* **83**, 513, 517 (2002).
- Masson, V. *et al.* Holocene climate variability in Antarctica based on 11 ice-core isotopic record. *Quat. Res.* **54**, 348–358 (2000).
- Jouzel, J. *et al.* Temperature reconstruction from Antarctic ice cores. *J. Geophys. Res.* (in the press).
- Krinner, G., Genthon, C. & Jouzel, J. GCM analysis of local influences on ice core δ signals. *Geophys. Res. Lett.* **24**, 2825–2828 (1997).
- Delaygue, G., Jouzel, J., Masson, V., Koster, R. D. & Bard, E. Validity of the isotopic thermometer in central Antarctica: limited impact of glacial precipitation seasonality and moisture origin. *Geophys. Res. Lett.* **27**, 2677–2680 (2000).
- Hoffmann, G., Masson, V. & Jouzel, J. Stable water isotopes in atmospheric General Circulation Models. *Hydrol. Process.* **14**, 1385–1406 (2000).
- Stenni, B. *et al.* An oceanic cold reversal during the last deglaciation. *Science* **293**, 2074–2077 (2001).
- Bender, M., Malaize, B., Orchado, J., Sowers, T. & Jouzel, J. *Mechanisms of Global Climate Change at Millennial Timescales* Vol. 112 (eds Clark, P. U., Webb, R. S. & Keigwin, L. D.) 149–164 (Geophysical Monograph, Washington DC, 1999).
- Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the Last Glacial Period. *Science* **291**, 109–112 (2001).
- Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
- Delaygue, G., Masson, V., Jouzel, J., Koster, R. D. & Healy, C. The origin of the Antarctic precipitation: a modelling approach. *Tellus* **27**, 19–36 (2000).
- Werner, M., Heimann, M. & Hoffmann, G. Isotopic composition and origin of polar precipitation in present and glacial climate simulations. *Tellus* **53**, 53–71 (2001).
- Yoshida, N. *et al.* Deuterium excess of past 320 kyr recorded in Antarctic Dome-Fuji ice core. *Int. Symp. on Dome Fuji Ice Core and Related Topics*, 20 (27–28 February 2001, Tokyo) (NIPR, Tokyo, 2001).
- Lorius, C. *et al.* A 150,000-year climatic record from Antarctic ice. *Nature* **316**, 591–596 (1985).
- Henderson, G. M. & Slowey, N. C. Evidence from U-Th dating against Northern Hemisphere forcing of the last deglaciation. *Nature* **404**, 61–66 (2000).
- Landwehr, J. M. & Winograd, I. J. Dating the Vostok ice core record by importing the Devils Hole chronology. *J. Geophys. Res.* **D 106**, 31853–31861 (2001).
- Delmotte, M. *Enregistrements Climatiques à Law-Dome: Variabilité pour les Périodes Récentes et pour la Déglaciation*. (Thesis, Université Joseph Fourier, Grenoble, 1997).
- Waelbroeck, C. *et al.* Sea level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002).

Acknowledgements We thank the Dome F drilling team and all participants in field work, ice sampling and isotopic measurements. We acknowledge the National Institute of Polar Research (Tokyo) for the logistic support. The project is supported in Japan by a Grant-in-Aid for Scientific Research on Priority Area (B) and, in France, by PNEDC (Programme National d'Etudes de la Dynamique du Climat) and by the Pole-Ocean-Pole (POP) programme of the European Communities. We thank R. Uemura and F. Vimeux for access to excess data, and N. Caillon, V. Masson, D. Schrag, B. Stauffer, R. Uemura, F. Vimeux and C. Waelbroeck for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.J. (e-mail: jouzel@lsce.saclay.cea.fr).

Reduced mixing from the breaking of internal waves in equatorial waters

Michael C. Gregg, Thomas B. Sanford & David P. Winkel

Applied Physics Laboratory, College of Fishery and Ocean Sciences, University of Washington, Seattle, Washington 98105, USA

In the oceans, heat, salt and nutrients are redistributed much more easily within water masses of uniform density than across surfaces separating waters of different densities. But the magnitude and distribution of mixing across density surfaces are also important for the Earth's climate as well as the concentrations of organisms¹. Most of this mixing occurs where internal waves break, overturning the density stratification of the ocean and creating patches of turbulence. Predictions of the rate at which internal waves dissipate^{2,3} were confirmed earlier at mid-latitudes^{4,5}. Here we present observations of temperature and velocity fluctuations in the Pacific and Atlantic oceans between 42° N and 2° S to extend that result to equatorial regions. We find a strong latitude dependence of dissipation in accordance with the predictions³. In our observations, dissipation rates and accompanying mixing across density surfaces near the Equator are less than 10% of those at mid-latitudes for a similar background of internal waves. Reduced mixing close to the Equator will have to be taken into account in numerical simulations of ocean dynamics—for example, in climate change experiments.

Internal waves are generated in the ocean when its stratification is disturbed. Unlike surface waves, which propagate horizontally because they are confined to the air–sea interface, internal waves propagate vertically as well as horizontally, making flow disturbances at the surface or bottom felt in the interior. Garrett and Munk⁶ and Munk⁷ modelled internal waves in terms of a wavenumber–frequency spectrum; they assumed that the wave field is constant in

time, and results from the superposition of many uncorrelated waves having frequencies between the buoyancy frequency N , determined by the stratification, and the Coriolis frequency, $f = 2\Omega_{\text{Earth}} \sin(\theta)$, where Ω_{Earth} is the Earth's rotation rate and θ is the latitude. The Garrett and Munk (GM) frequency spectrum is 'red', with most of the energy very close to f . Integrating components of the GM energy spectrum over the entire range of frequency and wavenumber yields average characteristics of the wave field, such as the net vertical displacement ζ and horizontal wave speed u . Subsequent observations show that the GM model represents the background state of the internal wave field, but some situations are significantly more energetic.

Numerical simulations of energy transfer by interactions within the internal wave field show a net flux towards small spatial scales that increases the shear variance $(\partial u/\partial z)^2$ until it overcomes the stratification and the waves break^{2,3}. When the wave field is statistically steady, or varies slowly in time, the rate at which wave breaking dissipates energy approximately equals the rate at which the energy is transferred from large to small scales. This equivalence allows the dissipation rate ε to be expressed in terms of internal wave parameters. To check their numerical simulations, Henyey, Wright and Flatté³ formulated an analytic model based on the Doppler shifting of evolving test waves by a background wave field. The Doppler shifting produces a net energy flux toward smaller scales—that is, higher wavenumbers—and ultimate breaking. The functional dependence of ε (which is expressed in units of W kg^{-1}) associated with this flux is the product of two terms:

$$\varepsilon = \varepsilon_{30^\circ}(N, \Phi_{\text{shear}}(m), \Phi_{\text{strain}}(m)) \times L(\theta, N) \quad (1)$$

The first term, ε at the 30° GM model reference latitude, depends on stratification (via N) and on internal wave characteristics (in terms of spectra of vertical shear, $\Phi_{\text{shear}}(m)$, and of vertical strain, $\Phi_{\text{strain}}(m)$, where m is the vertical wavenumber in cycles per metre). Strain fluctuations are variations of the vertical separations between density surfaces, that is, fluctuations of $\partial \zeta/\partial z$. Previous measurements⁴ verified the shear and N dependence in ε_{30° , and subsequent observations⁵ and numerical simulations⁸ confirmed the strain dependence. The second term contains the latitude (θ) dependence:

$$L(\theta, N) = \frac{f \cosh^{-1}(N/f)}{f_{30^\circ} \cosh^{-1}(N_0/f_{30^\circ})} \quad (2)$$

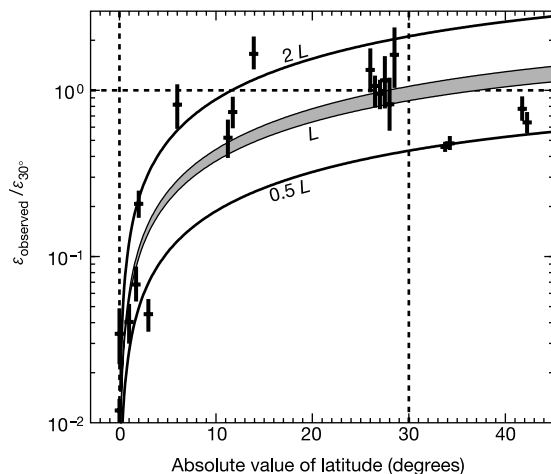


Figure 1 Reduction of dissipation rates produced by breaking internal waves near the Equator. The predicted latitude effect L is unity at 30° because the standard internal wave spectrum, Garrett and Munk⁶, and the calculations based on it are referenced to that latitude. Ratios of average $\varepsilon_{\text{observed}}$ to ε_{30° are plotted as short horizontal bars, and upper and lower 95% confidence limits are shown by extents of the vertical lines. Multiple values at the same latitude θ come from different depths or locations. The shaded curve gives the predicted $L(\theta, N)$ for the range of N encountered. Because the N dependence is weak, the height of the shaded band is small. Upper and lower solid lines are twice and one-half the prediction, and approximately bound the scatter in the data. In spite of the scatter, the observations confirm the predicted sharp cut-off of dissipation at low latitude.

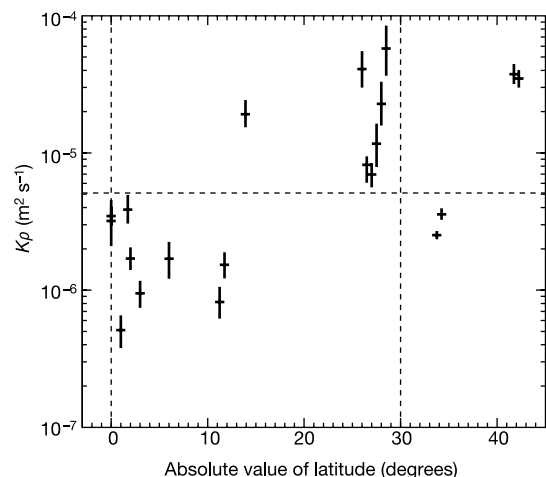


Figure 2 Diapycnal diffusivities. K_ρ was calculated using observed dissipation rates ε and stratification N^2 in equation (2). The horizontal reference line is the value of K_ρ at 30° latitude when internal waves are at the level of the Garrett and Munk spectrum. Owing to the latitude effect on ε produced by internal waves, intense internal waves observed near the Equator produced only modest K_ρ . For reference, the molecular diffusivity of heat in water is $1.4 \times 10^{-7} \text{ m}^2 \text{ s}^{-1}$.

and is unity at N_0 and 30° . L varies weakly with N , but is independent of internal wave characteristics. The latitude effect arises because the rate at which the evolving waves are Doppler shifted is proportional to the ratio between their vertical and horizontal length scales⁹. This ratio, in turn, depends on their dominant frequency, which is always slightly greater than f and thus varies with latitude.

Our measurements testing L were taken with the Multi-scale Profiler¹⁰ (MSP), a free-fall instrument that resolves velocity and temperature fluctuations from the length of the profile, usually 1,000 m, to centimetre scales of turbulent dissipation. N^2 and strain spectra are obtained from profiles of temperature and salinity. Electromagnetic sensors and an acoustic current meter produce velocity profiles yielding shear with metre-scale resolution. These allow us to evaluate ε_{30° . The turbulence parameter $\varepsilon_{\text{observed}}$ is obtained independently by measuring centimetre-scale velocity fluctuations with small sensitive probes (airfoils).

In Fig. 1, ratios of $\varepsilon_{\text{observed}}$ to ε_{30° are plotted for sets of profiles between the equator and 42° N. Solid curves are L and its multiples. The data scatter by two-fold about L , masking the latitude effect at mid-latitudes⁴, but not at low latitudes, where dissipation rates are sharply cut off. Near and on the Equator, $\varepsilon_{\text{observed}}$ is only 1–6% of mid-latitude ratios, confirming the prediction.

Most numerical models express diapycnal mixing in terms of an eddy coefficient K_ρ . When the average dissipation rate nearly balances the rate at which breaking internal waves produce turbulence, the mixing can be estimated simply¹¹ as:

$$K_\rho \approx 0.2\varepsilon/N^2 \quad (3)$$

where K_ρ is in units of $\text{m}^2 \text{s}^{-1}$. Application of this local estimate to the large-scale mean circulation is *ad hoc*¹², but comparisons of observations like ours with the rate of thickening of tracer clouds injected into the thermocline¹³ and into abyssal^{14,15} waters confirm the estimates. For our observations, K_ρ is markedly smaller at low latitudes than at mid-latitudes (Fig. 2). The smallest values are only a few times larger than the molecular diffusivity of heat in these waters, $1.4 \times 10^{-7} \text{ m}^2 \text{s}^{-1}$, demonstrating that at some times and places diapycnal mixing is negligible in the equatorial ocean.

To appreciate fully the effect of the low-latitude cut-off of internal wave mixing, it is necessary to examine the intensity of the internal wave field. The most important term in ε_{30° is the square of the ratio of the observed shear spectrum to the shear variance predicted by the GM model: this term varies from 0.32 to 1,140 (Fig. 3). In view

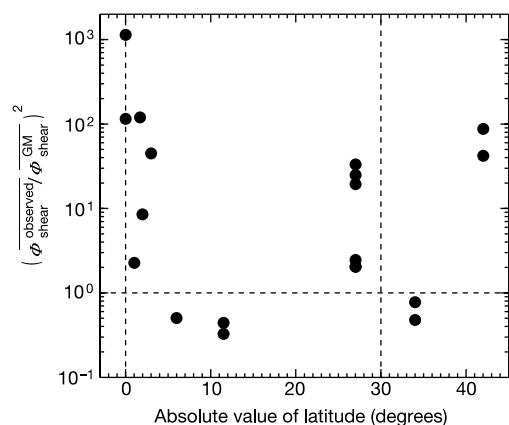


Figure 3 Effect of internal wave intensity on dissipation. The most important scaling term in equation (4) is $(0.1/k_c)^2$ which is plotted here in its more physical form as $(\phi_{\text{observed}} / \phi_{\text{shear}}^{\text{GM}})^2$. The horizontal dotted line at unity is the level for the GM model internal wave spectrum at 30° latitude. Some of the low-latitude measurements reveal very intense internal waves close to and on the Equator.

Table 1 Multi-scale profile

Latitude	Longitude	Expedition	Refs	Profiles
0°	156° E	COARE3	19	34
0°	140° W	Tropic Heat 2	20	69
1° N	140° W	Tropic Heat 2	20	7
1.7° S	156° E	COARE3	19	28
2° N	140° W	Tropic Heat 2	20	5
3° N	156° E	COARE3	19	12
6° N	140° W	Tropic Heat 2	20	3
11.5° N	134.8° W	Tropic Heat 2	20	5
13.9° N	59° W	C-SALT edge	21, 22	4
27° N	79.4° W	Florida Straits strn 5&6	23	10
34° N	127° W	PATCHEX	4, 24	27
42° N	126° W	PATCHEX north	4	5

Locations, expedition names with citations of relevant publications, and number of profiles of the data used in this comparison.

of such a large range, the two-fold scatter in the scaled dissipation rates is rather small. The three largest spectral ratios are on and close to the Equator. Were it not for the sharp latitudinal cut-off near the Equator, this trio would have produced the largest K_ρ in our data.

The agreement between observations and theory at low latitudes is better than expected (F. Henyey, personal communication), because it is unlikely that the GM model should be accurate close to the Equator and linear approximations made to obtain the dissipation model may not hold. For instance, the shear/strain correction accounts for only one of the ways in which a wave field can depart from the GM model, and more radical departures may occur where f is very small. We, therefore, interpret Fig. 1 as confirmation that the wave-wave interactions leading to ultimate breaking, energy dissipation, and diapycnal mixing do indeed vary with latitude, and that the effect is most important at low latitudes. The result is sufficiently robust to be used in numerical models, but internal wave measurements that are more detailed and of longer duration than ours are needed before we can be confident that we fully understand the mechanisms involved. For example, small but finite dissipation rates were observed on the Equator where no dissipation is predicted. They may indicate the importance of nonlinearities not considered in the prediction, but they could also result from either small mean shears enhancing mixing above the levels produced solely by internal wave interactions, or from effects of horizontal changes in mean currents, which are also neglected in the prediction.

Because mixing matters in the ocean, it must be represented accurately in models. Mapping dissipation rates throughout the ocean is a daunting task that is made considerably easier if mixing can be estimated from internal wave parameters, stratification and latitude; internal wave fields appear to be horizontally coherent over much larger scales than mixing patches, and to vary more slowly. Consequently, if the production of mixing by internal waves is understood and quantified, the problem of mapping mixing can be transferred to larger spatial scales and slower timescales by mapping internal waves. This requires a large effort, but the task is feasible. The result, however, would not be adequate for predicting mixing in future oceans having altered stratification and forced by differing wind patterns. Developing realistic climate prediction also requires understanding of how internal waves are generated, so the global field can be modelled realistically. For example, the only mixing measurements repeated at the same place during different seasons showed hints of a seasonal dependence¹⁶, and numerical modelling is now showing how internal waves may vary in response to seasonal changes in wind stress¹⁷. A first attempt at a comprehensive model is under way (P. Muller, personal communication, 2002), but many more observations of internal waves and mixing will be needed to verify its predictions. The observations presented here confirm part of the input needed for comprehensive understanding and global modelling of internal waves and mixing. □

Methods

MSP data

Table 1 gives the positions and number of MSP profiles forming the station averages, and references to the expeditions that collected them. We used data from below the seasonal thermocline, where the GM model should apply, and where the mean shear was weak. Owing to the strong mean shear in the main thermocline near and on the Equator during Tropic Heat 2, only data from the lower thermocline were used. COARE3 data on the Equator were used between 400–640 m and 640–890 m because we found no significant mean shear. Mean shear was weak near the centre of the Straits of Florida at stations 5 and 6, which were averaged between 205–305, 305–405 and 405–505 m. The C-SALT profiles were taken just outside the thermohaline staircase.

Dissipation rates

Shear spectra of airfoil records were calculated over 2-m intervals, and integrated to estimate local dissipation rates along each profile. Histograms of each depth interval and station were examined to estimate the noise level for the ensemble. The noise levels varied from 5×10^{-11} to $8 \times 10^{-11} \text{ W kg}^{-1}$, probably owing to differing noise characteristics of the airfoils. Estimates at or below the noise level were set to zero on the assumption that no mixing occurred. This was consistent with the lack of identifiable density overturns in these intervals, but we could not always be certain owing to the half-metre vertical resolution attainable with the Sea-Bird temperature and salinity records. Averages and confidence limits were computed using bootstrap¹⁸ techniques. Zeroing noisy ε values made little difference in most averages, but it reduced the COARE3 deep average at 3°N by one-half, and the Tropic Heat 2 data at 11.5°N by two-thirds. Using the unconditional averages would put the former within and the latter above the plotted bounds of 0.5 and 2 times the prediction.

Predicting the dissipation rate

Dissipation rates were predicted by evaluating the first part of equation (1) expressed as:

$$\varepsilon_{30^\circ} = 6.73 \times 10^{-10} \left(\frac{N}{N_0} \right)^2 \left(\frac{0.1}{k_c} \right)^2 \left(\frac{1+1/R}{4/3} \right) \left(\frac{2}{R-1} \right)^{1/2} \quad (4)$$

where ε_{30° is in units of W kg^{-1} . This equation of is a slight modification of the formulation of ref. 5. N_0 is the GM model reference stratification $5.24 \times 10^{-3} \text{ s}^{-1}$, k_c is the high-wavenumber cut-off of the internal wave shear spectrum, and R is a dimensionless ratio defined below. To accommodate varied spectral shapes, the shear spectrum is integrated until the variance reaches $0.661 N^2$, that is, $\int_{k_0}^{k_c} \Phi_{\text{shear}}^{\text{observed}}(m) dm = 0.661 N^2$. The upper limit k_c is taken to be the high-wavenumber cut-off, and k_0 is the reciprocal of the length of record being analysed. The reference variance is obtained by integrating the white GM shear spectrum to its cut-off at 0.1 cycles per metre, that is $\int_{0.005}^{0.1} \Phi_{\text{shear}}^{\text{GM}}(m) dm = 0.661 N^2$. Consequently, at vertical wavenumbers less than k_c , the ratio of the average spectral levels is $\Phi_{\text{shear}}^{\text{observed}} / \Phi_{\text{shear}}^{\text{GM}} = 0.1/k_c$. The second term in equation (4) is the square of this ratio, and is plotted in Fig. 3. R is the average ratio of the shear variance normalized by N^2 to the strain variance, both integrated to k_c . Dependent on R , the third and fourth terms in equation (4) are corrections for variations in the ratio of total energy to horizontal kinetic energy and in the dominant frequency content, both relative to the GM model.

The latitude dependence, equation (2), is the analytic form in the theoretical model³ and has not been previously tested. As noted earlier, it arises because the rate at which the evolving waves are Doppler shifted depends on the ratio between their vertical and horizontal length scales. This is expressed by $\frac{f_{\text{horiz}}}{f_{\text{vert}}} = \left(\frac{\omega^2 - f^2}{N^2 - \omega^2} \right)^{1/2} = \frac{f}{N} \cosh^{-1} \left(\frac{N}{f} \right)$, where averaging is across the internal wave frequency range, $f \lesssim \omega \lesssim N$, and is weighted by the frequency content of the GM wave field. Comparison to the GM reference at 30° leads to equation (2). Because N/f is a large number, $\cosh^{-1}(N/f) \approx \ln(2N/f)$.

Received 19 July 2002; accepted 14 February 2003; doi:10.1038/nature01507.

1. National Science Foundation *Ocean Sciences at the New Millennium* (NSF, Washington DC, 2001).
2. McComas, C. H. & Muller, P. The dynamic balance of internal waves. *J. Phys. Oceanogr.* **11**, 970–986 (1981).
3. Henyey, F. S., Wright, J. & Flatte, S. M. Energy and action flow through the internal wave field: An eikonal approach. *J. Geophys. Res.* **91**, 8487–8495 (1986).
4. Gregg, M. C. Scaling turbulent dissipation in the thermocline. *J. Geophys. Res.* **94**, 9686–9698 (1989).
5. Polzin, K., Toole, J. M. & Schmitt, R. W. Finescale parameterization of turbulent dissipation. *J. Phys. Oceanogr.* **25**, 306–328 (1995).
6. Garrett, C. J. R. & Munk, W. H. Space-time scales of internal waves: A progress report. *J. Geophys. Res.* **80**, 291–297 (1975).
7. Munk, W. H. in *Evolution of Physical Oceanography* (eds Warren, B. A. & Wunsch, C.) 264–291 (MIT Press, Cambridge, Massachusetts, 1981).
8. Sun, H. & Kunze, E. Internal wave-wave interactions. Part II: Spectral energy transfer and turbulence production. *J. Phys. Oceanogr.* **29**, 2905–2919 (1999).
9. Henyey, H. F. in *Dynamics of Internal Gravity Waves in the Ocean: Proceedings, 'Aha Huliko'a Hawaiian Winter Workshop* (eds Muller, P. & Henderson, D.) 233–236 (Spec. publ. SOEST, Univ. Hawaii, 1991).
10. Winkel, D. P., Gregg, M. C. & Sanford, T. B. Resolving oceanic shear and velocity with the Multi-Scale Profiler. *J. Atmos. Ocean. Technol.* **13**, 1046–1072 (1996).
11. Osborn, T. R. Estimates of the local rate of vertical diffusion from dissipation measurements. *J. Phys. Oceanogr.* **20**, 83–89 (1980).
12. Davis, R. E. Diapycnal mixing in the ocean: The Osborn-Cox model. *J. Phys. Oceanogr.* **24**, 2560–2576 (1994).
13. Ledwell, J. R., Watson, A. J. & Law, C. S. Evidence for slow mixing across the pycnocline from an open-ocean tracer release experiment. *Nature* **364**, 701–703 (1993).
14. Toole, J. M., Polzin, K. L. & Schmitt, R. W. Estimates of diapycnal mixing in the abyssal ocean. *Science* **264**, 1120–1123 (1994).

15. Ledwell, J. R. *et al.* Evidence for enhanced mixing over rough topography in the abyssal ocean. *Nature* **403**, 179–182 (2000).
16. Gregg, M. C. Variations in the intensity of small-scale mixing in the main thermocline. *J. Phys. Oceanogr.* **7**, 436–454 (1977).
17. Nagasawa, M., Niwa, Y. & Hibiya, T. Spatial and temporal distribution of wind-induced internal wave energy available for deep water mixing in the North Pacific. *J. Geophys. Res.* **105**, 13933–13943 (2000).
18. Efron, G. & Gong, G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am. Stat.* **37**, 36–48 (1983).
19. Gregg, M. C., Winkel, D. P., Sanford, T. B. & Peters, H. Turbulence produced by internal waves in the oceanic thermocline at mid and low latitudes. *Dynam. Atmos. Oceans* **24**, 1–14 (1996).
20. Peters, H., Gregg, M. C. & Sanford, T. B. Equatorial and off-equatorial shear variability at 140°W . *J. Geophys. Res.* **96**, 16913–16928 (1991).
21. Schmitt, R. W., Perkins, H., Boyd, J. D. & Stalcup, M. C. C-SALT: an investigation of the thermohaline staircase in the western tropical North Atlantic. *Deep-Sea Res.* **34**, 1655–1665 (1987).
22. Gregg, M. C. & Sanford, T. B. Shear and turbulence in thermohaline staircases. *Deep-Sea Res.* **34**, 1689–1696 (1987).
23. Winkel, D. P., Gregg, M. C. & Sanford, T. B. Patterns of shear and turbulence across the Florida Current. *J. Phys. Oceanogr.* **32**, 3269–3285 (2002).
24. Gregg, M. C. & Sanford, T. B. The dependence of turbulent dissipation on stratification in a diffusively stable thermocline. *J. Geophys. Res.* **93**, 12381–12392 (1988).

Acknowledgements The Office of Naval Research and the National Science Foundation funded our instrumentation and data collection. A SECNAV/CNO fellowship in oceanography supported M.C.G. during the analysis and writing of this letter.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C.G. (e-mail: gregg@apl.washington.edu).

Cannibalism in the Madagascan dinosaur *Majungatholus atopus*

Raymond R. Rogers*, David W. Krause† & Kristina Curry Rogers‡

* Geology Department, Macalester College, 1600 Grand Avenue, St Paul, Minnesota 55105, USA

† Department of Anatomical Sciences, Stony Brook University, Stony Brook, New York 11794, USA

‡ Department of Paleontology, Science Museum of Minnesota, 120 West Kellogg Boulevard, St Paul, Minnesota 55102, USA

Many lines of evidence have been brought to bear on the question of theropod feeding ecology, including functional and physiological considerations, morphological constraints, taphonomic associations, and telling—although rare—indications of direct ingestion^{1–7}. Tooth marks of theropods, although rarely described and generally left unassigned to a particular taxon, can provide unique clues into predator–prey interaction⁸, and can also yield insights into the extent of carcass utilization^{9,10}. Here we describe a sample of tooth-marked dinosaur bone recovered from three well-documented localities in the Upper Cretaceous Maevarano Formation of Madagascar that provides insights into the feeding ecology of the abelisaurid theropod *Majungatholus atopus*¹¹. Intensely tooth-marked elements from multiple individuals show that *Majungatholus* defleshed dinosaur carcasses. Furthermore, *Majungatholus* clearly fed upon the remains of not only sauropods, but also conspecifics, and thus was a cannibal. Cannibalism is a common ecological strategy among extant carnivores, but until now the evidence in relation to carnivorous dinosaurs has been sparse and anecdotal.

Majungatholus atopus inhabited the plains of northwestern Madagascar during the Late Cretaceous. Its bones and teeth are found throughout the Maevarano Formation (Campanian–Maastrichtian), but are most abundant within the channel-belt deposits of the Anembalemba Member¹². Here, the remains of *Majungatholus* are commonly found in association with those of

other vertebrate taxa in spectacular 'bonebeds'¹³. For example, quarry MAD96-01 yielded a nearly complete skull, lower jaws, and associated postcranial elements of a single *Majungatholus* individual (Field Museum of Natural History FMNH PR2100)¹¹, as well as the remains of fish, turtles, crocodyliforms, sauropods, birds and mammals. At least 12 postcranial elements of *Majungatholus* preserved at this locality (including five chevrons, three neural arches, two transverse processes, one neural spine and one rib) exhibit tooth marks (Fig. 1). Three chevrons, one transverse process and the neural spine exhibit parallel sets of tooth marks. Two chevrons from this site exhibit drag marks from denticles on the tooth carinae.

Another bonebed in the Anembalemba Member, quarry MAD96-21, yielded disarticulated skull elements, the left ilium, and most of the precaudal axial column of a subadult *Majungatholus* (University of Antananarivo UA 8678) in association with rare frog, turtle, crocodyliform and sauropod elements. The bones of this subadult individual are again scored by numerous conspicuous tooth marks, and at least nine *Majungatholus* elements recovered from this site exhibit evidence of feeding by a fairly large and persistent vertebrate carnivore. The most frequently marked elements are ribs, although at least two neural arches from the site also exhibit tooth marks. Two ribs in the sample exhibit bite marks on both superficial and deep aspects of the element. Multiple sets of denticle drag marks accompany tooth marks on one rib (Fig. 2a, b).

Denticle drag marks on two specimens (FMNH PR2100 (chevron, field number 96313-31D, quarry MAD96-01) and UA 8678 (rib, field number 96300-C, quarry MAD96-21): Figs 1c and 2a, b)

were measured, and average 2.1 denticles mm^{-1} . Ten *Majungatholus* crowns from quarry MAD93-18 were selected at random for comparison. On average, the ten teeth are characterized by 2 denticles mm^{-1} , with a range of 1.9–2.5 denticles mm^{-1} . Measurements taken from the modified bones and the *Majungatholus* teeth are definitely comparable. The set of parallel tooth marks on specimen 96313-31D (Fig. 1c) was also measured to approximate intertooth spacing in the perpetrator's jaw. Tooth marks on this element are spaced 1.0–1.7 cm apart. Comparable intertooth spacing, with values ranging from 1.3 to 2.3 cm, characterizes the tooth-bearing elements of the *Majungatholus* specimen (FMNH PR2100) collected from MAD96-01: the slightly lower spacing values on the chevron are easily explained by the angle at which the tooth row intersected the bone. It should also be noted that *Majungatholus atopus* can exhibit an even pattern of tooth eruption, as exemplified by the right dentary of FMNH PR2100 (Fig. 2c). An even pattern of tooth eruption would be needed to generate the set of tooth marks on specimen 96313-31D, and several other bones in the sample. These observations (with the caveats that tooth eruption varies among individuals, and intertooth spacing varies with ontogeny), in combination with the denticle drag patterns described above, make a compelling case for *Majungatholus* feeding on *Majungatholus* carrion at both MAD96-01 and MAD96-21.

Additional evidence of feeding behaviour can be found in collections from quarry MAD93-18, which include the remains of the small theropod *Masiakasaurus knopfleri*¹⁴, juvenile and adult specimens of the sauropod *Rapetosaurus krausei*¹⁵, along with fish, turtles, snakes, crocodyliforms, birds, mammals, and the bones and

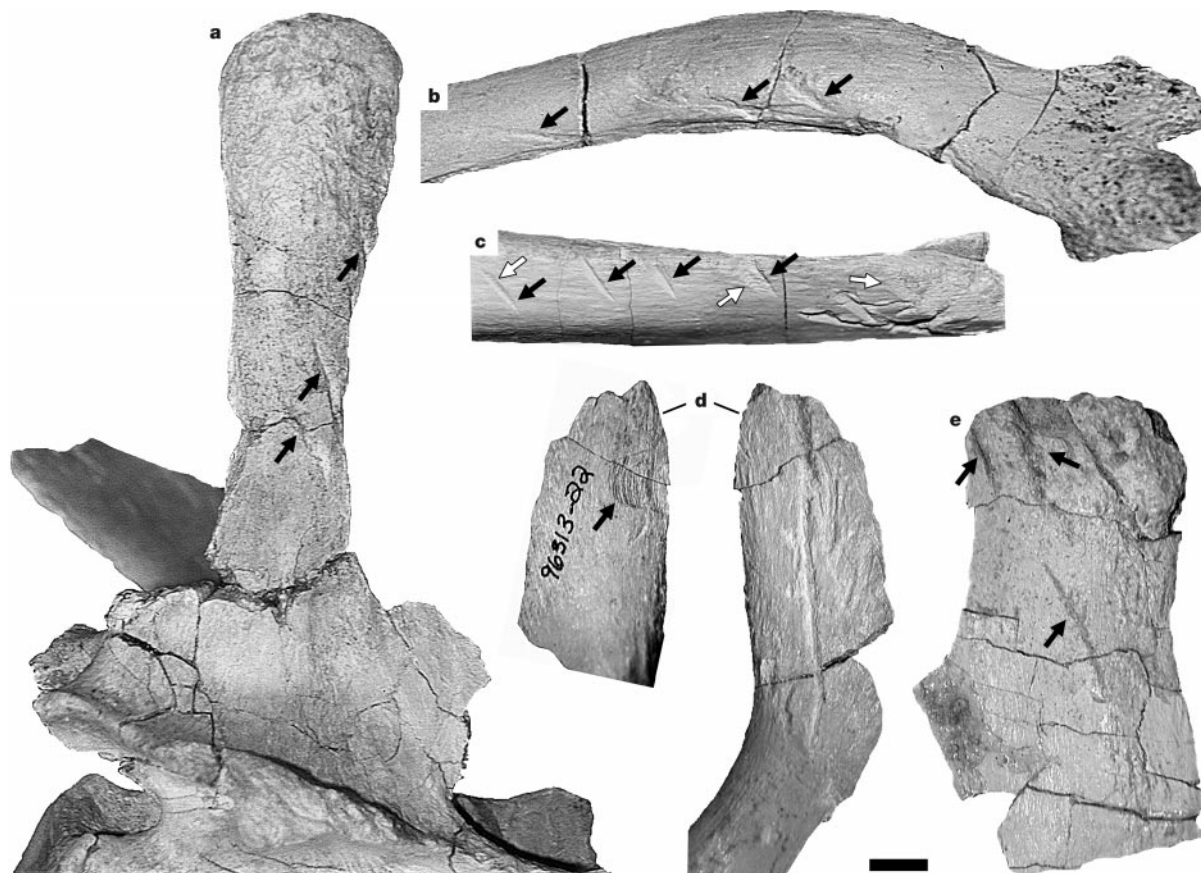


Figure 1 Representative sample of tooth-marked *Majungatholus atopus* bones (FMNH PR2100) from quarry MAD96-01. **a**, Caudal vertebra (field number 96313-14) with set of parallel tooth marks (arrows) on anterior edge of neural spine. **b**, Chevron (field number 96212-44C) with set of parallel tooth marks (arrows) on posterior margin. **c**, Chevron (field number 96313-31D) with set of four parallel tooth marks (black arrows), deep tooth

gouges, and denticle drag marks (white arrows). **d**, Chevron (field number 96313-22) gouged with denticle drag marks (right-lateral view, arrow) and deep tooth drag (left-lateral view). **e**, Transverse process (field number 96313-31) with multiple tooth marks (arrows). Scale bar, 1 cm.

teeth of *Majungatholus*. A large *Rapetosaurus* pubis from this locality exhibits a conspicuous concentration of sub-parallel furrows along the element's distal margin (Fig. 3). The size, shallow U-shaped morphology, and subparallel configuration of the traces are consistent with infliction by the premaxillary teeth of *Majungatholus*, which may have been processing the end of this element in an attempt to procure marrow and/or spongy bone. At least two additional bones from this site, a theropod rib and a sauropod chevron, also exhibit tooth marks.

The only other theropod known from the Maevarano Formation is *Masiakasaurus knopfleri*, whose small size precludes it from generating the tooth marks in question (Fig. 4). It is also possible to rule out the contemporaneous large crocodyliforms *Trematochamps*¹⁶ and *Mahajangasuchus*¹⁷, because their robust, conical teeth were too blunt, too irregularly spaced, too variable in height, too variable in orientation (especially *Mahajangasuchus*, whose anterior teeth were procumbent), and too variably positioned (buccolingually) relative to the longitudinal axis of the jaw (especially *Trematochamps*) to inflict the sets of narrow U-shaped grooves documented on the bones of *Majungatholus* (Figs 1 and 4). Of the medium- to large-bodied carnivores in the Maevarano Formation's vertebrate assemblage, only *Majungatholus atopus* possessed dentition capable of inflicting the suite of tooth marks documented in quarries MAD96-01, MAD96-21 and MAD93-18.

Data from three separate bonebeds indicate that *Majungatholus*

atopus regularly defleshed dinosaur carcasses. The main focus seems to have been along the axial column, although appendicular elements such as the sauropod pubis in quarry MAD93-18 also show evidence of concerted effort. There is also indication that *Majungatholus atopus* fed upon the remains of its own dead, and thus supplemented its diet via cannibalism. The evidence comes in the form of at least 21 tooth-marked elements derived from two *Majungatholus* individuals (FMNH PR2100 and UA 8678) preserved in two separate localities (quarries MAD96-01 and MAD96-21). Whether *Majungatholus* killed both individuals and thus practiced intraspecific predation, or opportunistically scavenged their remains, is unknown.

Cannibalism as an ecological strategy is not uncommon among extant animals. At least 14 species of mammalian carnivores kill and eat members of their own species, and numerous reptilian and avian taxa also practice cannibalism^{18–22}. In contrast, the evidence for cannibalism among dinosaurs is meagre. The most celebrated example of a dinosaur cannibal is the Triassic theropod *Coelophysis*

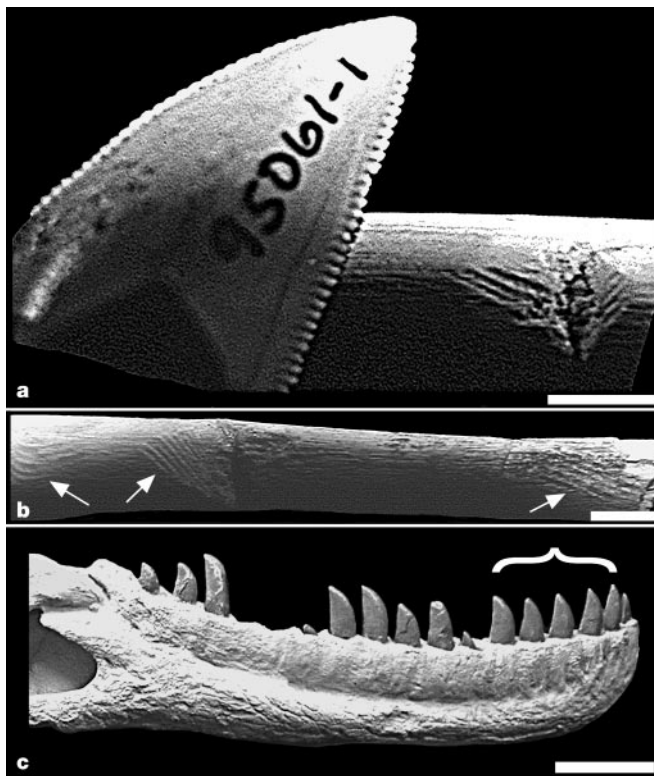


Figure 2 Physical evidence linking *Majungatholus atopus* to the tooth marks. **a, b**, *Majungatholus atopus* rib (UA 8678, field number 96300-C) from quarry MAD96-21 that exhibits multiple sets of denticle drag marks (both views are different regions of the same rib). The chisel-like morphology and spacing of the denticles on the *Majungatholus* tooth shown in **a** are comparable to the denticle patterns on modified bones. Teeth of *Majungatholus* occasionally exhibit damaged and/or worn denticles, which presumably reflects contact with bone. Scale bars, 5 mm. **c**, The right dentary of *Majungatholus atopus* (FMNH PR2100) collected from MAD96-01 exemplifies the even pattern of tooth eruption (beneath bracket) required to generate the parallel sets of tooth marks documented on several bones in the sample. The intertooth spacing of *Majungatholus* is also consistent with the pattern of tooth marks. Scale bar, 5 cm.



Figure 3 *Rapetosaurus krausei* pubis (FMNH PR2255) from quarry MAD93-18 with evidence of peripheral feeding along the element's distal margin. The size, shallow U-shaped morphology, and subparallel configuration of the furrows are consistent with infliction by premaxillary teeth of *Majungatholus atopus*. These traces are morphologically distinct from trample marks, which tend to be much smaller, and to occur in sets of parallel striations. Scale bar, 1 cm.

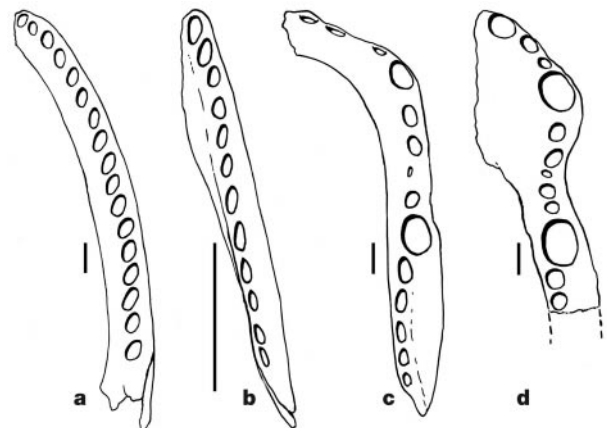


Figure 4 Occlusal views of right dentaries of selected carnivorous taxa in the Maevarano Formation. Taxa shown include *Majungatholus atopus* (**a**), *Masiakasaurus knopfleri* (**b**), *Mahajangasuchus insignis* (**c**) and *Trematochampsia oblita* (**d**). All scale bars, 2 cm. The theropod *Masiakasaurus* was too small to inflict the suite of tooth marks in question. The crocodyliforms *Mahajangasuchus* and *Trematochampsia* both possessed dentition that effectively rules them out as well: their conical teeth were too blunt, too irregularly spaced, too variable in height, and too variable in orientation to inflict the tooth marks on the bones of *Majungatholus atopus* and *Rapetosaurus krausei*.

bauri^{23,24}, although a recent reappraisal of the evidence suggests that claims of cannibalistic feeding by *Coelophysis* may be unsubstantiated. Juvenile *Coelophysis* bones traditionally interpreted as an ingested meal may lie stratigraphically at least partially below the rib cage of the proposed cannibal (as opposed to within), and the volume of the purportedly ingested material may exceed reasonable estimates of stomach capacity²⁵. Cannibalism has also been tentatively proposed for Late Cretaceous tyrannosaurids on the basis of the occurrence of tooth marks on tyrannosaurid bone. However, multiple tyrannosaurid taxa with morphologically indistinguishable teeth occur in the same unit that yields the tooth-marked bones, so the evidence is equivocal⁸. Other claims of dinosaur cannibalism are largely anecdotal²⁶. Thus, *Majungatholus atopus* at the very least joins *Coelophysis bauri* as a dinosaur cannibal, and (on present evidence) may in fact be the only theropod dinosaur with demonstrable cannibalistic tendencies. □

Received 14 November 2002; accepted 17 February 2003; doi:10.1038/nature01532.

- Chin, K., Tokaryk, T. T., Erickson, G. M. & Calk, L. C. A king-sized theropod coprolite. *Nature* **393**, 680–682 (1998).
- Erickson, G. M. *et al.* Bite-force estimation for *Tyrannosaurus rex* from tooth-marked bones. *Nature* **382**, 706–708 (1996).
- Rayfield, E. J. *et al.* Cranial design and function in a large theropod dinosaur. *Nature* **409**, 1033–1037 (2001).
- Farlow, J. O. A consideration of the trophic dynamics of a Late Cretaceous large-dinosaur community (Oldman Formation). *Ecology* **57**, 841–857 (1976).
- Farlow, J. O., Brinkman, D. L., Abler, W. L. & Currie, P. J. Size, shape and serration density of theropod dinosaur lateral teeth. *Mod. Geol.* **16**, 161–198 (1991).
- Maxwell, W. D. & Ostrom, J. H. Taphonomy and paleobiological implications of *Tenontosaurus-Deinonychus* associations. *J. Vert. Paleontol.* **15**, 707–712 (1995).
- Varricchio, D. J. Gut contents from a Cretaceous tyrannosaurid: implications for theropod digestive tracts. *J. Paleontol.* **75**, 401–406 (2001).
- Jacobsen, A. R. Feeding behaviour of carnivorous dinosaurs as determined by tooth marks on dinosaur bones. *Hist. Biol.* **13**, 17–26 (1988).
- Fiorillo, A. R. Prey bone utilization by predatory dinosaurs. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **88**, 157–166 (1991).
- Hunt, A. P., Meyer, C. A., Lockley, M. G. & Lucas, S. G. Archaeology, toothmarks and sauropod dinosaur taphonomy. *Gaia* **10**, 225–231 (1994).
- Sampson, S. D. *et al.* Predatory dinosaur remains from Madagascar: Implications for the Cretaceous biogeography of Gondwana. *Science* **280**, 1048–1051 (1998).
- Rogers, R. R., Hartman, J. H. & Krause, D. W. Stratigraphic analysis of Upper Cretaceous rocks in the Mahajanga Basin, Madagascar: Implications for ancient and modern faunas. *J. Geol.* **108**, 275–301 (2000).
- Krause, D. W. *et al.* The Late Cretaceous vertebrate fauna of Madagascar: Implications for Gondwanan paleobiogeography. *GSA Today* **9**, 1–7 (1999).
- Carrano, M. T., Sampson, S. D. & Forster, C. A. The osteology of *Masiakasaurus knopfleri*, a small abelisauroid (Dinosauria: Theropoda) from the Late Cretaceous of Madagascar. *J. Vert. Paleontol.* **22**, 510–534 (2002).
- Curry Rogers, K. A. & Forster, C. A. The last of the dinosaur titans: a new sauropod from Madagascar. *Nature* **412**, 530–534 (2001).
- Buffetaut, E. & Taquet, P. Un nouveau Crocodylien méso-suchien dans le Campanien de Madagascar: *Trematocampa obliata*, n. sp. *Bull. Soc. Géol. Fr.* **21**, 183–188 (1979).
- Buckley, G. A. & Brochu, C. A. in *Cretaceous Fossil Vertebrates* (ed. Unwin, D.) 149–175 (Palaeontological Assoc., London, 1999).
- Bertram, B. Social factors influencing reproduction in wild lions. *J. Zool.* **177**, 463–482 (1975).
- Polis, G. A. The evolution and dynamics of intraspecific predation. *Annu. Rev. Ecol. Syst.* **12**, 225–251 (1981).
- Auffenberg, W. *The Behavioral Ecology of the Komodo Monitor* 406 (Univ. Press of Florida, Gainesville, 1981).
- Fox, L. R. Cannibalism in natural populations. *Annu. Rev. Ecol. Syst.* **6**, 87–106 (1975).
- Polis, G. A., Myers, C. A. & Hess, W. P. A survey of intraspecific predation within the class Mammalia. *Mammal Rev.* **14**, 187–198 (1984).
- Colbert, E. H. The Triassic dinosaur *Coelophysis*. *Mus. North. Arizona Bull. Ser.* **57**, 1–160 (1989).
- Colbert, E. H. *The Little Dinosaur of Ghost Ranch* 250 (Columbia Univ. Press, New York, 1995).
- Gay, R. J. The myth of cannibalism in *Coelophysis bauri*. *J. Vert. Paleontol.* **22**, 57A (2002).
- Cooper, M. R. The prosauropod dinosaur *Massospondylus carinatus* Owen from Zimbabwe: Its biology, mode of life, and phylogenetic significance. *Occ. Pap. Nat. Mus. Rhodesia* **6**, 689–840 (1981).

Acknowledgements We are grateful to the villagers of Berivotra, the staff of the Institute for the Conservation of Tropical Environments, and the faculty of the University of Antananarivo for logistical support, and all field crew members of the Mahajanga Basin Project for their dedicated efforts. We thank L. Betti-Nash for providing artwork for Fig. 4, and K. Chin, G. Erickson, J. Holstein, B. Koralessky, P. O'Connor, S. Sampson, W. Simpson and J. Thole for comments and discussions. This work was supported by NSF, National Geographic Society and Macalester College.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.R.R. (e-mail: rogers@macalester.edu).

Conservation of total synaptic weight through balanced synaptic depression and potentiation

Sébastien Royer & Denis Paré

Center for Molecular & Behavioral Neuroscience, Rutgers State University, 197 University Avenue, Newark, New Jersey 07102, USA

Memory is believed to depend on activity-dependent changes in the strength of synapses¹. In part, this view is based on evidence that the efficacy of synapses can be enhanced or depressed depending on the timing of pre- and postsynaptic activity^{2–5}. However, when such plastic synapses are incorporated into neural network models, stability problems may develop because the potentiation or depression of synapses increases the likelihood that they will be further strengthened or weakened⁶. Here we report biological evidence for a homeostatic mechanism that reconciles the apparently opposite requirements of plasticity and stability. We show that, in intercalated neurons of the amygdala, activity-dependent potentiation or depression of particular glutamatergic inputs leads to opposite changes in the strength of inputs ending at other dendritic sites. As a result, little change in total synaptic weight occurs, even though the relative strength of inputs is modified. Furthermore, hetero- but not homosynaptic alterations are blocked by intracellular dialysis of drugs that prevent Ca^{2+} release from intracellular stores. Thus, in intercalated neurons at least, inverse heterosynaptic plasticity tends to compensate for homosynaptic long-term potentiation and depression, thus stabilizing total synaptic weight.

To control runaway synapses, many neural network models use normalization algorithms where activity modifies the relative strength of inputs while conserving total synaptic weights or overall neuronal activity⁷. Although these algorithms do not pretend to biophysical realism, they imply the existence of an intracellular signalling system that would render synapses 'aware' of each other or of mean neuronal activity. Here, we tested whether synaptic normalization is operative when particular inputs to a given neuron are subjected to stimulation protocols producing long-term depression (LTD) and long-term potentiation (LTP). For methodological reasons that will become clear below, these experiments were performed in intercalated (ITC) neurons of the amygdala⁸. ITC cells are spiny GABA (γ -aminobutyric acid) neurons located between the basolateral amygdala (BLA) and the central nucleus (Fig. 1A, grey ovals). They receive topographically organized glutamatergic inputs from the BLA.

To activate different groups of BLA axons converging on single ITC cells, an array of closely spaced ($\sim 150 \mu\text{m}$) stimulating electrodes (Fig. 1A, black dots) was positioned in the BLA. Below, stimulation sites are given arbitrary numbers that increase as their position shifts lateromedially. To assess the selectivity of this stimulation method, we tested whether responses evoked in ITC cells ($n = 9$) from different BLA sites occluded each other. Stimuli were delivered at each BLA site independently (Fig. 1Ba, top) or two at a time (Fig. 1Ba, bottom). The amplitude of excitatory postsynaptic currents (EPSCs) evoked by paired stimuli, normalized to that observed following summation of responses evoked by single stimuli, was plotted as a function of the interval between paired stimulation sites (Fig. 1Bb). Response occlusion was seen only with neighbouring sites (paired t -test, $P < 0.05$), suggesting that this stimulation method was relatively selective.

The BLA sites that evoked the largest responses were always located at the same mediolateral level as the recorded soma. This suggested that there is a correspondence between the mediolateral

position of BLA stimulation sites, and where activated axons end on ITC cells. To test this, excitatory postsynaptic potentials (EPSPs) evoked by different BLA sites (Fig. 1Ca, b) were recorded in nine ITC cells. Data obtained from different cells were aligned to the BLA site evoking the largest response (termed site 0 in Fig. 1) and averaged. Average EPSP rise times (open circles) and amplitudes (filled triangles), respectively, increased and decreased as the distance between the stimulation site and recorded soma increased (Fig. 1Cc). Moreover, plotting EPSP rise times against their amplitudes (Fig. 1Cd) revealed an inverse correlation between these two variables ($r = 0.9$; $P < 0.05$), suggesting that there is an orderly spatial relationship between the position of BLA stimulation sites and where activated axons end on ITC cells⁹. Thus, ITC cells

are ideally suited to study the heterosynaptic effects of activity-dependent plasticity.

Low- and high-frequency BLA stimuli (LFS, HFS) are known to, respectively, produce homosynaptic NMDA (*N*-methyl-D-aspartate)-dependent LTD and LTP in ITC cells⁸. To determine whether LFS and HFS also produce inverse heterosynaptic modifications, electrical stimuli were delivered sequentially at all effective BLA sites before (15 min) and after (40 min) applying LFS or HFS at one site (termed site 0 in Figs 2–4).

LFS produced a significant depression of responses evoked from the induction site in 14 out of 17 cells (*t*-tests, $P < 0.05$). In contrast, responses elicited from many heterosynaptic sites were potentiated (LFS effect versus stimulation sites, $F = 7.9$, d.f. = 14, $P < 0.05$). The effect of LFS on the response profile of a representative cell is shown in Fig. 2Aa, b (control, filled squares; 26–40 min post-LFS, open squares) and Fig. 2Ba. Figure 2Bb depicts the time course of changes in normalized response amplitude for all sites that reached significance (–4, –3, –2 and 5; *t*-tests, $P < 0.05$, stepwise Bonferroni correction¹⁰). Note that the heterosynaptic potentiation was already apparent at the end of LFS. The same results were obtained averaging results from all cells with a significant homosynaptic LTD (Fig. 2Ac, Bc; LFS effect versus stimulation sites, $F = 3.84$, d.f. = 4, $P < 0.05$). In Fig. 2Ac, the heterosynaptic potentiation was significant at sites 2, 4 and 5 (paired *t*-tests,

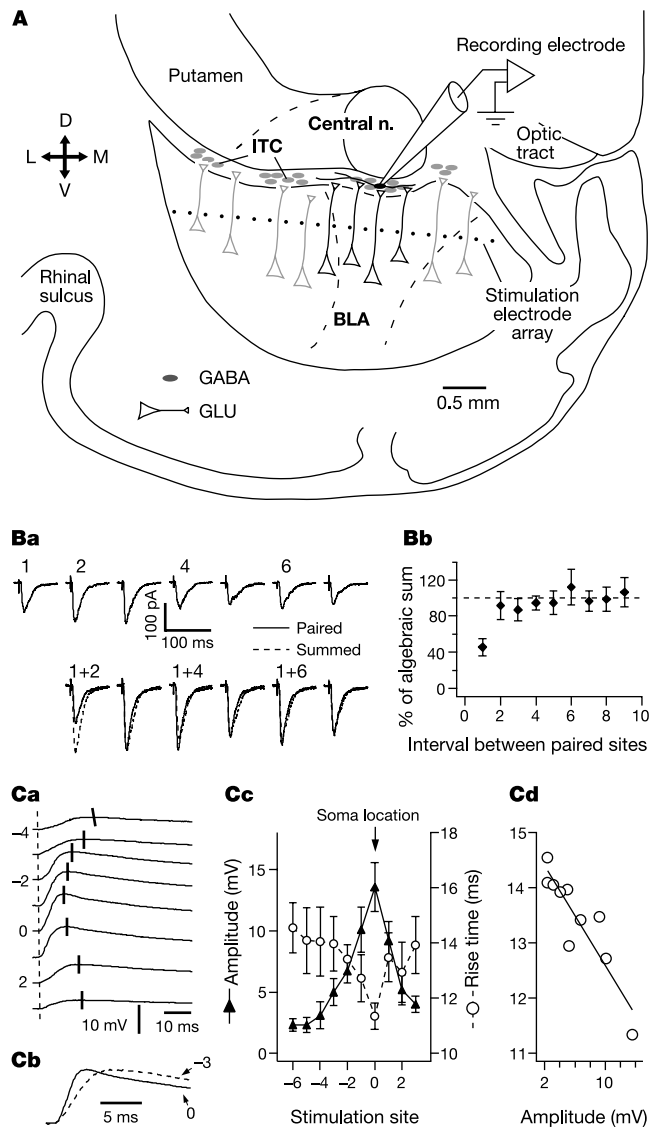


Figure 1 Experimental approach. **A**, Coronal slice of the amygdala. Stimulating electrode array (dots) used to activate BLA afferents to ITC cells (grey ovals). **B**, Occlusion test. **Ba**, BLA stimuli applied one (top row) or two (bottom row) at a time. Numbers indicate relative position of stimulation sites (lateromedially). **Bb**, EPSCs evoked by paired stimuli, normalized to amplitude of summed responses, plotted as a function of interval between stimulation sites. **Ca**, Examples of evoked EPSPs. **Cb**, EPSP 0 and –3 after scaling to match amplitude. **Cc**, Plot of EPSP amplitude and rise time versus stimulation site. **Cd**, Plot of EPSP amplitude versus rise time. Central n., central nucleus of the amygdala; D, dorsal; GLU, glutamatergic; L, lateral; M, medial; V, ventral.

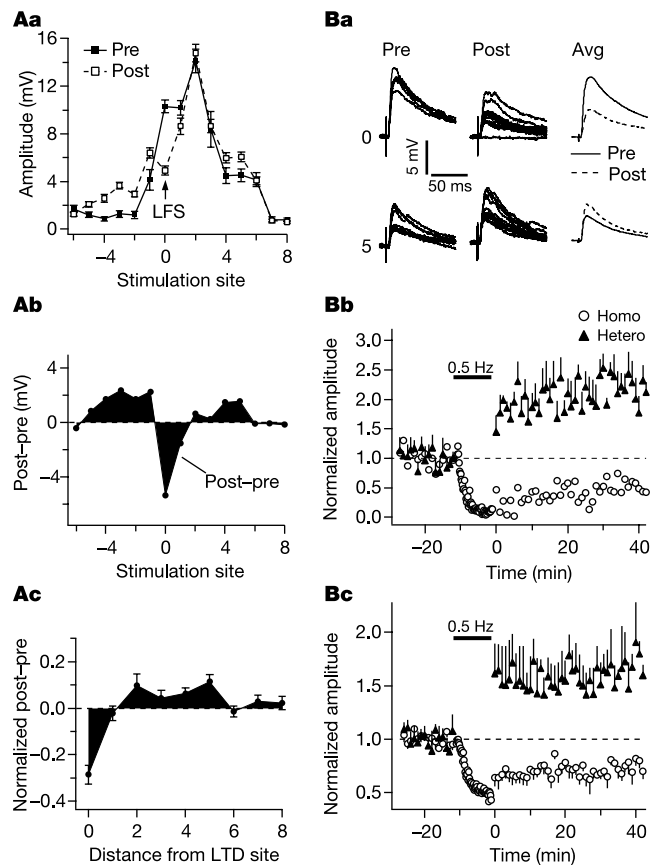


Figure 2 LTD induction produces heterosynaptic LTP. **Aa**, EPSP amplitudes as a function of stimulation site before/after LFS at site 0. **Ab**, Difference between pre- and post-LFS response profiles for same cell. **Ac**, As in **Ab** but average of all cells ($n = 14$). **Ba**, Examples of EPSPs evoked from sites 0 and 5 before/after LTD induction. Same cell as in **Aa**, **b**. **Bb**, Time course of changes in response amplitude for same cell. **Bc**, As in **Bb** but for all cells combined. In **Bb**, **c**, empty circles represent responses evoked from site 0. Filled triangles represent responses evoked from heterosynaptic sites that showed significant potentiation. Avg, average.

$P < 0.05$). Similar results were obtained irrespective of the medio-lateral location of the induction site relative to the recorded cell.

Application of the NMDA glutamate receptor antagonist AP-5 ($D(-)$ -2-amino-5-phosphonopentanoic acid; $100 \mu\text{M}$) before (3 min) and during LTD induction blocked the homo- and heterosynaptic effects of LFS (LFS effect versus stimulation sites, $F = 0.11$, d.f. = 5, $P > 0.05$, $n = 9$; see Supplementary Information). Similarly, when recorded cells were dialysed with the calcium chelator BAPTA ($1,2$ -bis(2-aminophenoxy)ethane- N,N,N',N' -tetraacetic acid; 50 mM), LFS did not change response amplitudes at homo- and heterosynaptic sites (LFS effect versus stimulation sites, $F = 0.27$, d.f. = 5, $P > 0.05$, $n = 7$; see Supplementary Information).

If inverse heterosynaptic changes constitute a mechanism for stabilizing total synaptic weights, then LTP induction should produce heterosynaptic depression. HFS produced homosynaptic LTP in 13 out of 27 cells (t -tests, $P < 0.05$). Heterosynaptic depression was seen only in cells with homosynaptic LTP induction. The response profile of a representative cell before and after HFS is shown in Fig. 3Aa, b (stimulation sites versus HFS effect, $F = 12.58$, d.f. = 10, $P < 0.05$) and Fig. 3Ba. Post-hoc t -tests revealed that responses evoked from sites 0 and 1 were significantly enhanced, whereas those elicited from sites 4 and 6 were depressed ($P < 0.05$).

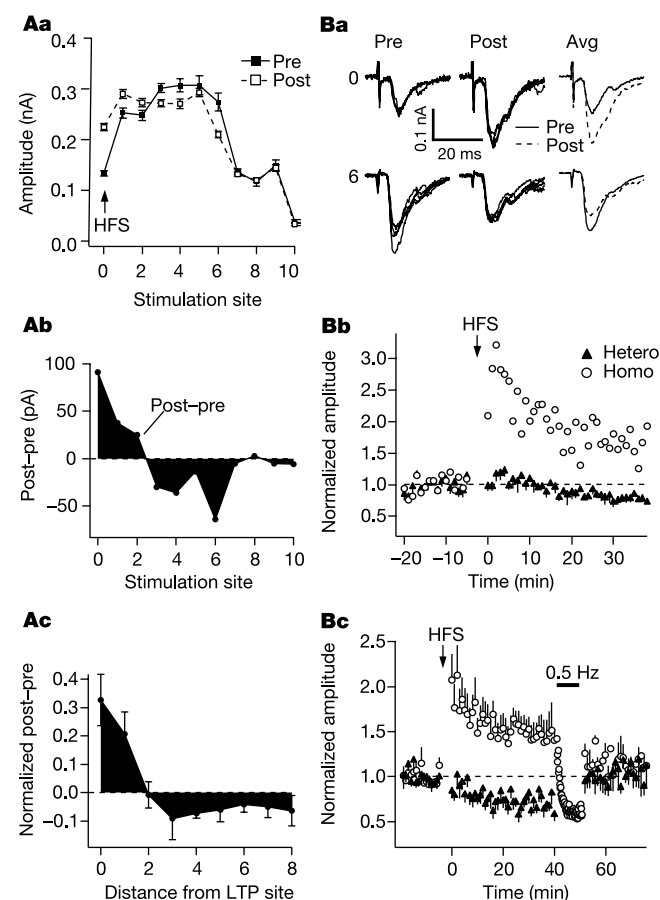


Figure 3 LTP induction produces heterosynaptic LTD. **Aa**, EPSC amplitudes as a function of stimulation site before/after HFS at site 0. **Ab**, Difference between pre- and post-LFS response profiles for same cell. **Ac**, As in **Ab** but average of all cells ($n = 13$). **Ba**, Examples of EPSCs evoked from sites 0 and 4 before/after LTD induction. Same cell as in **Aa**, **b**. **Bb**, Time course of changes in response amplitude for same cell. **Bc**, As in **Bb** but for all cells combined. In **Bb**, **c**, open circles represent responses evoked from site 0. Filled triangles represent responses evoked from heterosynaptic sites that showed significant depression.

In contrast to the rapid appearance of heterosynaptic effects in LTD experiments, inverse heterosynaptic changes developed gradually after LTP induction (Fig. 3Bb).

Similar results were obtained after averaging results from cells with a significant homosynaptic LTP (Fig. 3Ac, HFS effects versus stimulation sites, $F = 2.53$, d.f. = 5, $P < 0.05$). Note that responses evoked from heterosynaptic sites adjacent to site 0 were potentiated, whereas those elicited by distant ones (sites 3–8) were depressed (t -tests, $P < 0.05$). The appearance of LTP at sites contiguous to site 0 might reflect the imperfect selectivity of our LTP induction protocol.

Homosynaptic LTP and heterosynaptic LTD did not develop when AP-5 ($100 \mu\text{M}$) was applied before (3 min) and during HFS ($n = 13$; HFS effect versus stimulation sites, $F = 0.04$, d.f. = 5, $P > 0.05$; see Supplementary Information). Moreover, intracellular dialysis of ITC cells with BAPTA (50 mM , $n = 7$) blocked the homo- and heterosynaptic effects of HFS (HFS effect versus stimulation sites, $F = 0.15$, d.f. = 5, $P > 0.05$; see Supplementary Information).

To determine whether heterosynaptic depression reflected a time-dependent phenomenon unrelated to LTP induction, we tested whether it could be reversed. To this end, five ITC neurons with successful LTP induction were subjected to LFS of BLA site 0, 40 min

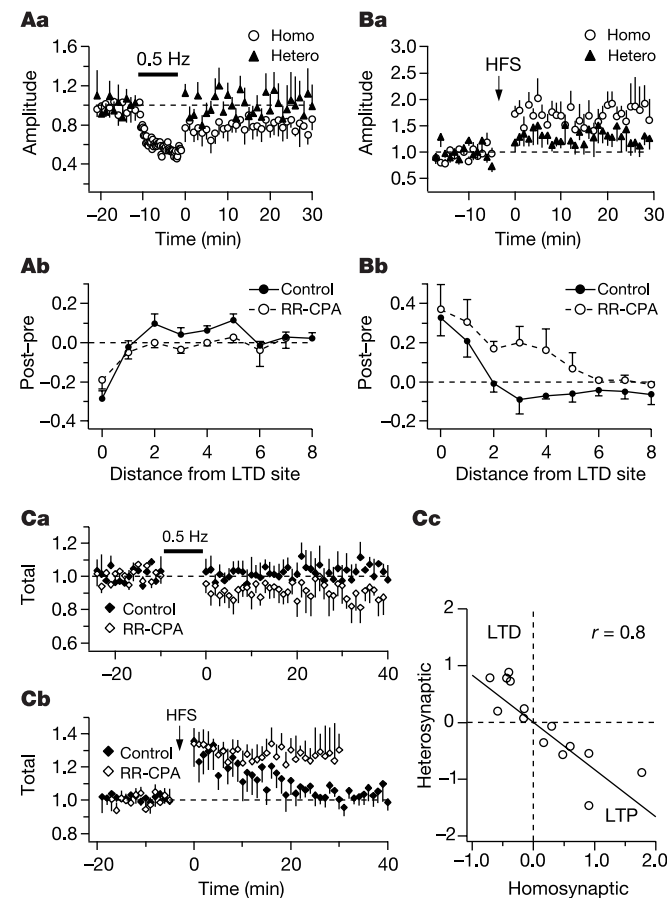


Figure 4 Ca^{2+} release from intracellular stores is required for the conservation of total synaptic weight. Effects of ruthenium red (RR) and cyclopiazonic acid (CPA) on homo- and heterosynaptic changes produced by LFS (**A**, **Ca**) and HFS (**B**, **Cb**). **(Aa, Ba)** Time course of changes in response amplitudes. **(Ab, Bb)** Response profile in control (filled circles) and RR/CPA experiments (empty circles). **Ca, b**, For each tested cells, responses evoked from all BLA sites were summed, normalized and used to compute a population average. **Cc**, Correlation between homo- and heterosynaptic plasticity in control conditions for a subset of cells recorded in the middle of the stimulating electrode array.

post-HFS (Fig. 3Bc). This treatment not only produced a homosynaptic depotentiation (open circles), but abolished heterosynaptic alterations (filled triangles, *t*-tests, $P > 0.05$). Moreover, repetitive supra-threshold (0.2 nA; 2 ms) current pulses (four trains of ten pulses at 100 Hz) in the absence of BLA stimuli ($n = 5$) did not affect the amplitude of BLA-evoked responses (see Supplementary Information).

A critical issue is to determine whether the above phenomena depended on plastic events that occurred in the BLA or at BLA–ITC synapses. We reasoned that the former possibility would seem unlikely if manipulations that affected only the recorded cell abolished hetero- but not homosynaptic changes. To examine this, we tested intracellular dialysis of drugs known to interfere with Ca^{2+} release from intracellular stores. This idea was based on evidence that Ca^{2+} entering through NMDA receptors can evoke Ca^{2+} -induced Ca^{2+} release^{11,12}, and that propagating waves of such release regulate heterosynaptic changes in other cell types¹³. Thus, cyclopiazonic acid ($n = 10$; 1 μM), a Ca^{2+} -ATPase inhibitor¹⁴, or ruthenium red ($n = 15$; 20 μM), an antagonist of ryanodine receptors¹⁵, were added to the pipette solution. Results obtained with both drugs will be pooled below, because they yielded qualitatively identical results (see Supplementary Information).

Cyclopiazonic acid and ruthenium red had no effect on control response amplitudes ($n = 25$), but they increased the likelihood that HFS would produce homosynaptic LTP (8 out of 9 cells; Fisher test, $P < 0.05$). In contrast, the incidence of successful homosynaptic LTD was not significantly modified by cyclopiazonic acid and ruthenium red (10 of 16 cells; Fisher test, $P > 0.05$). Below, we consider only data from cells with successful induction of homosynaptic LTD or LTP.

Response modifications produced by LFS and HFS in experiments with ruthenium red or cyclopiazonic acid differed significantly from those seen in control conditions (LFS, $F = 17.6$, d.f. = 4, $P < 0.05$; HFS, $F = 12.8$, d.f. = 4, $P < 0.05$). Yet, the amount of homosynaptic depression (Fig. 4A) or potentiation (Fig. 4B) was the same (*t*-tests, $P > 0.05$). The main difference resided in the heterosynaptic effects of LFS (Fig. 4A) and HFS (Fig. 4B). The heterosynaptic potentiation seen at sites 2–5 in control LFS experiments was abolished (Fig. 4Ab; $P > 0.05$). In LTP experiments, the heterosynaptic depression seen at sites 3–5 in control conditions was replaced by a potentiation ($P < 0.05$).

The normalization hypothesis stipulates that activity-dependent synaptic plasticity is achieved by relative changes in synaptic strengths with no overall modification in total synaptic weight. Because measuring total synaptic weight is experimentally impossible, we tested this idea by examining how summed responses were affected by LTD and LTP induction. To this end, amplitudes of responses evoked from all effective BLA sites were added. Results from each cell were then normalized to control values, and a population average was computed. Whereas LFS had no effect on summed responses in control conditions (Fig. 4Ca, filled diamonds), it produced a persistent depression with ruthenium red or cyclopiazonic acid (Fig. 4Ca, open symbols). In control LTP experiments, HFS first increased summed responses, but they returned to initial levels in ~20 min (Fig. 4Cb, filled symbols). However, this time-dependent decay did not take place with ruthenium red or cyclopiazonic acid (Fig. 4Cb, open symbols). Finally, a significant inverse correlation was found between homo- and heterosynaptic plasticity in control conditions (Fig. 4Cc; $r = -0.8$; $P < 0.05$).

Our results suggest that activity-dependent enhancement or depression of particular inputs to ITC neurons is accompanied by inverse modifications at heterosynaptic sites, thus contributing to the stabilization of total synaptic weight. Although heterosynaptic LTD has been observed in other structures following LTP^{16–19} or LTD induction^{13,20}, to the best of our knowledge, there has been no report of heterosynaptic LTP following LTD induction.

Consistent with studies in CA1 pyramidal neurons^{21–23}, our results suggest that Ca^{2+} entry through NMDA receptors is essential for homosynaptic LTD and LTP induction in ITC cells. This is indicated by the prevention of homosynaptic LTP and LTD by application of AP-5 during induction, or by dialysis of ITC cells with BAPTA. However, these treatments also abolished the heterosynaptic effects of LFS and HFS, suggesting that homo- and heterosynaptic changes are causally related, or that they constitute parallel processes triggered by the same factor.

Surprisingly, drugs interfering with Ca^{2+} -induced Ca^{2+} release abolished the heterosynaptic, but not the homosynaptic, effects of HFS and LFS. It may seem paradoxical that Ca^{2+} -induced Ca^{2+} release could lead to heterosynaptic LTD and LTP. At present, it is believed that homosynaptic LTP requires large rises in intracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$), whereas lower $[\text{Ca}^{2+}]_i$ rises lead to homosynaptic LTD. Although our results are consistent with this view at homosynaptic sites, we also show that Ca^{2+} -induced Ca^{2+} release can produce either heterosynaptic LTD or LTP, depending on homosynaptic stimulation parameters. The crucial question then is how do unstimulated inputs detect the stimulation frequency at the stimulated pathway? We suggest that $[\text{Ca}^{2+}]_i$ transients evoked at heterosynaptic sites vary depending on the stimulation frequency at homosynaptic sites. In cases of homosynaptic LTP, it is likely that, besides Ca^{2+} -induced Ca^{2+} release, propagating action potentials were also evoked, leading to the activation of voltage-gated Ca^{2+} channels. In contrast, when homosynaptic LTD was induced, spiking did not occur and, presumably, only Ca^{2+} -induced Ca^{2+} release was evoked. This suggests that an entirely different logic links local variations in $[\text{Ca}^{2+}]_i$ and plasticity in the absence of synaptic activity. Indeed, our results imply that a higher $[\text{Ca}^{2+}]_i$ leads to depression of inactive inputs, whereas $[\text{Ca}^{2+}]_i$ transients of lower magnitude, presumably as produced by Ca^{2+} -induced Ca^{2+} release alone, produce a potentiation of inactive synapses.

In this context, it should be mentioned that besides differences in $[\text{Ca}^{2+}]_i$, HFS and LFS probably produced differences in the subcellular distribution of Ca^{2+} ions. We surmise that these variations ultimately affect the phosphorylation/dephosphorylation^{24,25} or the insertion/removal rate of glutamate receptor subunits²⁶, leading to spatially heterogeneous effects on glutamate-mediated synaptic transmission. However, other possibilities, such as the differential modulation of transmitter release or of postsynaptic voltage-dependent conductances, should not be overlooked.

The homeostatic mechanism that we report here adds to the growing list of processes implicated in the regulation of plastic synapses⁶. Although inverse homo- versus heterosynaptic plasticity is reminiscent of synaptic scaling²⁷, these two mechanisms operate on different timescales (minutes versus days). At present, it is unclear whether inverse homo- versus heterosynaptic plasticity could stabilize total synaptic weight if ITC neurons were challenged with more realistic, spatially distributed activation patterns. Finally, this homeostatic mechanism might coexist with other regulatory processes, such as spike-timing-based hebbian plasticity^{2–5,28} or synaptic redistribution^{29,30}.

In conclusion, our results suggest that, in ITC neurons, LTP and LTD are not local events engaging a limited group of synapses, but are cell-wide phenomena in which the distribution of synaptic weights is constantly updated, each synapse being affected by the fate of other, distant inputs. □

Methods

Coronal slices of the amygdala were obtained from guinea-pigs (3–5 weeks old) as previously described⁸. After a 1-h recovery period, slices were transferred to a chamber perfused with a warm (32 °C), oxygenated artificial cerebrospinal fluid containing (in mM) 126 NaCl, 2.5 KCl, 1.25 NaH_2PO_4 , 1 MgCl_2 , 2 CaCl_2 , 26 NaHCO_3 and 10 glucose. Current- or voltage-clamp recordings were obtained under visual control with borosilicate pipettes filled with a solution containing (in mM) 130 potassium gluconate, 10 N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid, 10 KCl, 2 MgCl_2 , 2 ATP-Mg and 0.2 GTP-tris(hydroxy-methyl)aminomethane. pH was adjusted to 7.2 and osmolarity to

280–290 mS. This Letter includes only results from neurons with a resting potential negative to -70 mV and input resistance >400 M Ω . Moreover, these recordings remained stable for >1 h with a low (<10 M Ω) and invariant series resistance ($<10\%$ variation). Picrotoxin (100 μ M) was present at all times. Homosynaptic LTD was induced with LFS of BLA inputs at 0.5 Hz for 10 min. Homosynaptic LTP was induced with HFS paired to postsynaptic depolarization. HFS consisted of four series of ten trains separated by 0.3 s, each train consisting of ten shocks at 100 Hz. Postsynaptic depolarization was achieved by applying short (2 ms) depolarizing current pulses (0.2 nA) timed so that BLA-evoked EPSPs would occur just before or during current-evoked spikes. The stimulation frequency in the control and post-treatment epochs was adjusted so that all effective BLA sites could be stimulated in one minute. Statistical tests always involved comparisons between 15 responses from control period and the post-LFS or HFS epoch (acquired 26 – 40 min after LTD or LTP induction). Statistical significance of the changes was assessed by computing two-way repeated-measures analysis of variance (ANOVA) and post-hoc t -tests with stepwise modification of the significance level ($P < 0.05$) for multiple comparisons¹⁰.

Received 31 December 2002; accepted 21 February 2003; doi:10.1038/nature01530.

- Martin, S. J., Grimwood, P. D. & Morris, R. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23**, 649–711 (2000).
- Markram, H., Lübke, J., Frotscher, M. & Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**, 213–215 (1997).
- Bi, G. Q. & Poo, M. M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
- Debanne, D., Gähwiler, B. H. & Thompson, S. M. Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures. *J. Physiol. (Lond.)* **507**, 237–247 (1998).
- Feldman, D. E. Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron* **27**, 45–56 (2000).
- Abbott, L. F. & Nelson, S. B. Synaptic plasticity: taming the beast. *Nature Neurosci.* **3**, 1178–1183 (2000).
- Arbib, M. A. *The Handbook of Brain Theory and Neural Networks* (MIT Press, Cambridge, Massachusetts, 1995).
- Royer, S. & Paré, D. Bidirectional synaptic plasticity in intercalated amygdala neurons and the extinction of conditioned fear responses. *Neuroscience* **115**, 455–462 (2002).
- Rall, W. in *Methods in Neuronal Modeling: From Synapses to Networks* (eds Koch, C. & Segev, I.) 9–62 (MIT Press, Cambridge, Massachusetts, 1989).
- Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
- Emptage, N., Bliss, T. Y. & Fine, A. Single synaptic events evoke NMDA receptor-mediated release of calcium from internal stores in hippocampal dendritic spines. *Neuron* **22**, 115–124 (1999).
- Liang, Y., Yuan, L. L., Johnston, D. & Gray, R. Calcium signaling at single mossy fiber presynaptic terminals in the rat hippocampus. *J. Neurophysiol.* **87**, 1132–1137 (2002).
- Nishiyama, M., Hong, K., Mikoshiba, K., Poo, M. M. & Kato, K. Calcium stores regulate the polarity and input specificity of synaptic modifications. *Nature* **408**, 584–588 (2000).
- Demaurex, N., Lew, D. P. & Krause, K.-H. Cyclopiazonic acid depletes intracellular Ca^{2+} stores and activates an influx pathway for divalent cations in HL-60 cells. *J. Biol. Chem.* **267**, 2318–2324 (1992).
- Xu, L., Tripathy, A., Pasek, D. A. & Meissner, G. Ruthenium red modifies the cardiac and skeletal muscle Ca^{2+} release channels (ryanodine receptors) by multiple mechanisms. *J. Biol. Chem.* **274**, 32680–32691 (1999).
- Lynch, G. S., Dunwiddie, T. & Gribkoff, V. Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature* **266**, 737–739 (1977).
- Abraham, W. C. & Wickens, J. R. Heterosynaptic long-term depression is facilitated by blockade of inhibition in area CA1 of the hippocampus. *Brain Res.* **546**, 336–340 (1991).
- Scanziani, M., Malenka, R. C. & Nicoll, R. A. Role of intercellular interactions in heterosynaptic long-term depression. *Nature* **380**, 446–450 (1996).
- Coussens, C. M. & Teyler, T. J. Long-term potentiation induces synaptic plasticity at nontetanized adjacent synapses. *Learn. Mem.* **3**, 106–114 (1996).
- Muller, D., Hefft, S. & Figurov, A. Heterosynaptic interactions between LTP and LTD in CA1 hippocampal slices. *Neuron* **14**, 599–605 (1995).
- Malenka, R. C. Synaptic plasticity in the hippocampus: LTP and LTD. *Cell* **78**, 535–538 (1994).
- Bliss, T. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Bear, M. F. & Abraham, W. C. Long-term depression in hippocampus. *Annu. Rev. Neurosci.* **19**, 437–462 (1996).
- Soderling, T. R. & Derkack, V. A. Postsynaptic protein phosphorylation and LTP. *Trends Neurosci.* **23**, 75–80 (2000).
- Winder, D. G. & Sweatt, J. D. Roles of serine/threonine phosphatases in hippocampal synaptic plasticity. *Nature Rev. Neurosci.* **2**, 461–474 (2001).
- Carroll, R. C., Beattie, E. C., von Zastrow, M. & Malenka, R. C. Role of AMPA receptor endocytosis in synaptic plasticity. *Nature Rev. Neurosci.* **2**, 315–324 (2001).
- Turrigiano, G. G. Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends Neurosci.* **22**, 221–227 (1999).
- Song, S., Miller, K. D. & Abbott, L. F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neurosci.* **3**, 919–926 (2000).
- Markram, H. & Tsodyks, M. Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature* **382**, 807–810 (1996).
- O'Donovan, M. J. & Rinzel, J. Synaptic depression: a dynamic regulator of synaptic communication with varied functional roles. *Trends Neurosci.* **20**, 431–433 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This work was supported by the National Science Foundation, the Canadian Institutes of Health Research and the Center for Molecular and Behavioral Neuroscience of Rutgers University. We thank Dr E. J. Lang for comments on the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.P. (e-mail: pare@axon.rutgers.edu).

Protection against enteric salmonellosis in transgenic mice expressing a human intestinal defensin

Nita H. Salzman^{*†}, Dipankar Ghosh^{†‡}, Kenneth M. Huttner[§], Yvonne Paterson^{||} & Charles L. Bevins[‡]

^{*} Department of Pediatrics, Division of Gastroenterology, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA

[‡] Department of Immunology, Lerner Research Institute, The Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA

[§] Division of Neonatology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA

^{||} Department of Microbiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA

[†] These authors contributed equally to this work

Genetically encoded antibiotic peptides are evolutionarily ancient and widespread effector molecules of immune defence^{1–3}. Mammalian defensins, one subset of such peptides, have been implicated in the antimicrobial defence capacity of phagocytic leukocytes and various epithelial cells⁴, but direct evidence of the magnitude of their *in vivo* effects have not been clearly demonstrated. Paneth cells, specialized epithelia of the small intestinal crypt, secrete abundant α -defensins and other antimicrobial polypeptides^{5,6} including human defensin 5 (HD-5; also known as DEFA5)^{7–9}. Although antibiotic activity of HD-5 has been demonstrated *in vitro*^{9,10}, functional studies of HD-5 biology have been limited by the lack of *in vivo* models. To study the *in vivo* role of HD-5, we developed a transgenic mouse model using a 2.9-kilobase HD-5 minigene containing two HD-5 exons and 1.4 kilobases of 5'-flanking sequence. Here we show that HD-5 expression in these mice is specific to Paneth cells and reflects endogenous enteric defensin gene expression. The storage and processing of transgenic HD-5 also matches that observed in humans. HD-5 transgenic mice were markedly resistant to oral challenge with virulent *Salmonella typhimurium*. These findings provide support for a critical *in vivo* role of epithelial-derived defensins in mammalian host defence.

Paneth cells reside in invaginations of the wall of the intestine called crypts of Lieberkühn, and are distributed along the length of the small intestine but are most abundant in the jejunum and ileum^{6,11}. In addition to α -defensins (termed cryptidins in mice), Paneth cells synthesize and secrete other antimicrobial polypeptides, including lysozyme and secretory group II phospholipase A2 (refs 6, 11). In humans, Paneth cells express two α -defensins named HD-5 and HD-6 (refs 7, 12). The release by Paneth cells of an array of antimicrobials is thought to contribute to host defence of the small intestine by influencing the composition and controlling the numbers of microbes in the crypt and lumen^{6,11}; however, *in vivo* data supporting this hypothesis are limited¹³.

To investigate the *in vivo* biological role of HD-5, we have developed a transgenic model in which HD-5 is expressed in mouse small intestinal Paneth cells. A comparison of the nucleotide sequence of the two human genes expressed in Paneth cells, HD-5 and HD-6, reveals a marked and rather unusual pattern of similarity—the 5′-flanking regions have sequence similarity greater than that observed in the coding regions (Fig. 1a). Reasoning that the 5′-flanking region might contain tissue-specific promoter elements, we generated HD-5 transgenic FvB mice using a 2.9-kilobase (kb) genomic fragment containing this flanking DNA and the adjacent coding exons (Fig. 1a, red bracket). Both heterozygous and homozygous HD-5 transgenic mice had normal development, fertility, intestinal histology and no observable phenotype when housed in a specific pathogen-free environment. The 1661 line contained ten copies of the transgene per diploid genome equivalent and a second line (6571) contained eight copies. These lines were indistinguishable in our analysis of transgene expression and phenotype, and for consistency, data for line 1661 are presented here. The third line (6568) contained one copy of the transgene and HD-5 expression was at levels too low for

northern detection. Except for line 6568 line, which showed less HD-5 expression than would be expected, the expression of the transgene appears primarily dependent on copy number, not on site of integration.

HD-5 transgene expression was assessed by multi-tissue northern blot analysis (Fig. 1b). The pattern of HD-5 transgene expression in the small intestine reflected that of endogenous mouse and human enteric α -defensins with a greater abundance of message distally than proximally. We also compared the expression level of HD-5 messenger RNA with that of cryptdin 4 (Fig. 1c), as the latter is reported to have maximal expression in the distal small intestine¹⁴. We found that cryptdin 4 mRNA levels do not vary significantly between wild-type and transgenic mice, indicating that transgenic expression of HD-5 does not alter or interfere with cryptdin expression. Furthermore, the HD-5 mRNA levels in transgenic mice are comparable to expression of the endogenous cryptdin 4 in mice, and the level of transgenic HD-5 mRNA is similar to that observed in the human ileal mucosa (Fig. 1c). *In situ* hybridization analysis localized HD-5 mRNA expression to the murine crypt Paneth cells (Fig. 1d). Therefore, the HD-5 transgene, under the

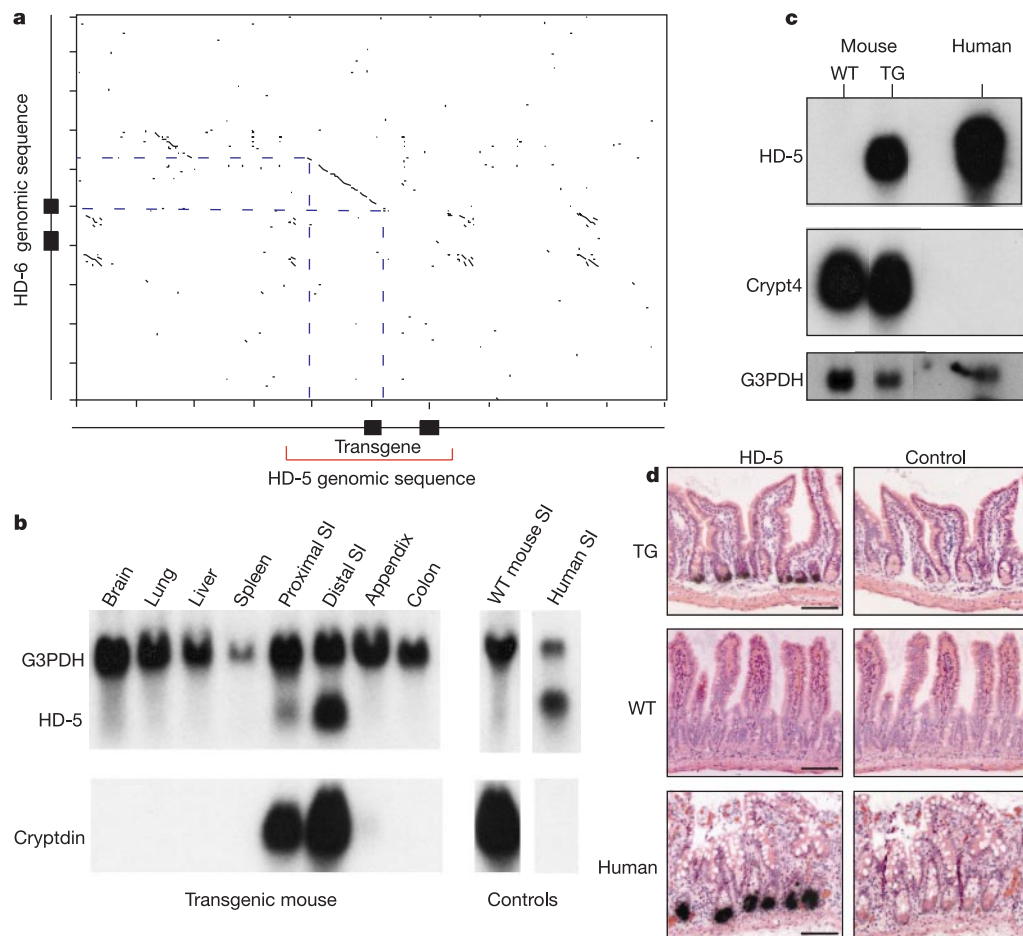


Figure 1 Germline transmission and characterization of HD-5 transgene expression in mice. **a**, Pustell analysis of HD-5 and HD-6 genomic sequences (window size 30 bp, similarity score $\geq 40\%$, hash value = 5). The flanking region immediately adjacent to exon 1 was similar in each gene (blue dashed lines). The 0.7-kb region in the HD-6 gene in the adjacent 5′-flanking sequence was also similar to a region in the HD-5 gene about 2 kb further upstream in its flanking sequence. A repetitive element in the intron and 3′-flanking region is repeated four times in the HD-5 genomic sequence. The transgenic genomic fragment, HG2-3e7, used in this study is indicated by the red bracket. Exons are indicated by solid boxes. **b**, Multi-tissue northern blot analysis of transgenic HD-5 mRNA expression. A G3PDH probe¹² was used as a control for RNA quantity and integrity. Human

and wild-type murine small intestinal RNA were used as controls of specific hybridization. **c**, Northern blot analysis of HD-5 transgenic and cryptdin 4 expression. RNA isolated from distal small intestine from transgenic (TG) and wild-type (WT) mice, and from human ileum were probed as described in **b**. **d**, *In situ* hybridization analysis. Expression of HD-5 mRNA by *in situ* hybridization using an antisense (AS) probe on sections of terminal ileum from a 35-day-old HD-5 transgenic mouse (TG), a wild-type age-matched control (WT) and an adult human specimen (Human) using methods as described¹². RNase A pre-treatment controls (right panels) and sense probes (data not shown) were negative for hybridization signal. Counter stain was haematoxylin and eosin. Scale bar, 40 μ m.

control of its own putative promoter, is appropriately expressed in a tissue- and cell-specific manner.

To characterize further HD-5 expression at the protein level, western blots were performed using acid-urea polyacrylamide gel electrophoresis (PAGE)^{9,15,16}. The blots detected HD-5 in the distal small intestine of transgenic mice, but not in samples with equal protein loading from the proximal small intestine or in tissues from wild-type littermate controls (Fig. 2a). The predominant form of HD-5 in the distal small intestine tissue of transgenic mice migrated similarly to HD-5 isolated from human intestinal tissue and similar to recombinant proHD-5(20–94) peptide (Fig. 2a, b). The lumen of transgenic mice contained a single fast-migrating form that migrated with recombinant mature HD-5 and HD-5 isolated from human ileal lumen (Fig. 2b). Recent studies have shown that HD-5 is stored in human Paneth cells as a propeptide^{9,17,18}, predominantly the 20–94 form, and that during or after secretion it is proteolytically processed to a 63–94 form recovered from the intestinal lumen⁹. To characterize further both the tissue and secreted forms, HD-5 peptides were isolated from small intestinal tissue washed free of luminal contents, and from the intestinal lumen after administration of an acetylcholine agonist to elicit Paneth cell secretion^{19,20}. The predominant HD-5 peptide isolated from the transgenic intestinal tissue was the 20–94 form, as determined by amino-terminal sequence and mass spectral analysis (the N-terminal sequence of the transgenic peptide was ESLQERA, identical to the human form⁹; observed m/z = 8103.9 daltons by matrix-assisted laser desorption/ionization (MALDI) mass spectroscopy, predicted m/z = 8102.9 daltons). The transgenic HD-5 peptide isolated from lumen was the 63–94 form, with N-terminal sequence (ATXYXRTG) identical to the secreted human form^{9,17}, and an observed mass of 3581.8 (predicted 3583.2). Thus, transgenic expression of HD-5 recapitulates expression of stored and secreted peptide as observed in humans. Using colony-forming unit (c.f.u.)-based assays, we also found that distal small-intestine-derived cationic proteins from transgenic mice have greater antibacterial activity against *Escherichia coli* compared with those from their wild-type littermates (data not shown).

We sought to determine whether the enhanced antimicrobial activity at this anatomically important site might provide protection against *S. typhimurium*, a murine enteric pathogen that is less sensitive *in vitro* to the bactericidal activity of endogenous murine cryptdins than to HD-5 (refs 9, 10, 21). HD-5 transgenic mice and age-matched FvB wild-type controls (4–5 weeks of age) were orally challenged with 1×10^8 c.f.u. virulent *S. typhimurium* (14028s). Bacteria surviving in the distal small intestine of the mice were plated on selective agar (Fig. 3a) and consistently were found to be ≥ 1 log-fold lower in the HD-5 transgenic mice compared with wild-type controls. When mice were challenged with higher numbers (1.5×10^9 c.f.u.), the phenotype conferred by transgene expression was more pronounced. At 6–24 h after inoculation the wild-type mice showed signs of progressive illness, with ruffled fur, hunched posture and diarrhoea (Fig. 3b). Mortality was 100% in wild-type control mice by 24 h after oral inoculation (Fig. 3c). In contrast, the HD-5 transgenic mice showed some initial signs of illness, but consistently recovered after 12 h with no morbidity or mortality (Fig. 3c). After this oral inoculation, numbers of recoverable *S. typhimurium* in the distal small intestine of transgenic and wild-type mice were determined (Fig. 3d). The total number of recoverable *S. typhimurium* colonies were 8×10^6 per 10 cm ileum in wild-type mice and 4×10^3 per 10 cm ileum in the HD-5 transgenic mice ($P < 0.001$, Student's *t*-test). Similarly, there were approximately 3 log-fold greater numbers of recoverable *S. typhimurium* in the faeces of wild-type control mice compared with HD-5 transgenic animals (data not shown). Thus, the *S. typhimurium* burden in the distal intestine is significantly lower in HD-5 transgenic compared with wild-type control mice, apparently as a result of HD-5 expression at this site. We suggest that HD-5 and

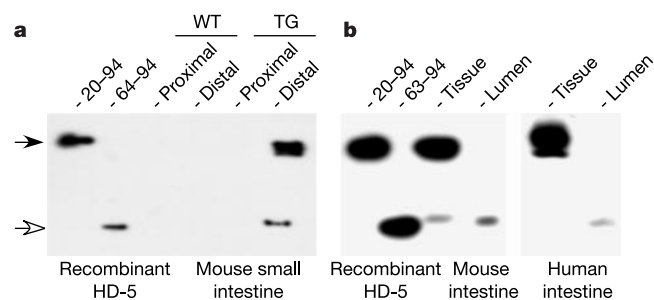


Figure 2 Western blot analysis of ileal HD-5 peptide. **a**, Western blot showing the expression pattern of HD-5 in transgenic mouse intestinal tissue. Tissue extracts from proximal and distal small intestinal tissues of wild-type and transgenic mice were loaded on 12.5% acid-urea PAGE, transferred on PVDF membranes and analysed for HD-5 immunoreactivity. Recombinant proHD-5(20–94) and mature HD-5(64–94) were used as standards (20 ng each lane). **b**, Western blot analysis of HD-5 peptide forms in small intestinal tissue and lumen of transgenic mice using recombinant HD-5 standards (left panel), and analysis of tissue and luminal forms of HD-5 in human small intestinal samples (right panel).

the aggregate of other mucosal host defence capacities of the mice combine to decrease the bacterial burden in the lumen and effectively contain the residual number of bacteria. In contrast, the significantly higher numbers of bacteria overwhelm the defence capacity in wild-type mice.

Additional studies of bacterial translocation support this conclusion. Homozygous HD-5 transgenic mice and age-matched controls were given sublethal doses of *S. typhimurium* orally. Spleens were examined for total *Salmonella* burden 3 days after inoculation (Fig. 3e). The HD-5 transgenic mice showed a tenfold reduction in *Salmonella* burden. If the effect of the HD-5 transgene expression on resistance to oral *Salmonella* were mediated in the intestinal lumen, bacterial challenge by alternative routes of administration would not show a difference when comparing transgenic with wild-type mice. We challenged groups of wild-type and transgenic mice by intraperitoneal inoculation of 10^4 – 10^6 c.f.u. *S. typhimurium*. Survival of transgenic and wild-type control mice was indistinguishable for each of the three levels of inocula (Fig. 3f). As data from an independent line of HD-5 transgenic mice (6571) show a highly similar phenotype, a significant effect of chromosomal insertion site of the transgene in this model system is excluded. Together, these studies are consistent with the hypothesis that the phenotype of the transgenic mice is a result of enhanced *Salmonella* killing within the intestinal lumen.

The HD-5 transgenic model has provided a means of investigating human defensin expression and function. The HD-5 transgenic mice have revealed that the promoter elements sufficient for Paneth cell expression of HD-5 are located within the 2.9-kb genomic fragment comprising the HD-5 transgene. The relative strength of the promoter elements within the transgene are probably attenuated as compared with those of the endogenous Paneth cell defensin genes, considering that the transgenic mice harbour approximately ten copies of the transgene yet yield expression levels comparable to endogenous cryptdin 4 in mouse intestine and HD-5 in human intestine. HD-5 transgenic peptide biosynthesis and processing is similar to that in humans, supporting the idea that this is a useful model to study the *in vivo* role of this human defensin. Mouse cryptdins are processed by a matrix metalloprotease called matrilysin 7 (MMP-7)^{13,22}, and mice rendered deficient in MMP-7 expression by targeted homologous recombination do not process cryptdins and are more susceptible to *Salmonella* challenge¹³. The processed form of HD-5 recovered from the lumen of transgenic mice, with proteolytic cleavage on the carboxyl side of arginine 62, is identical to the form recovered from the

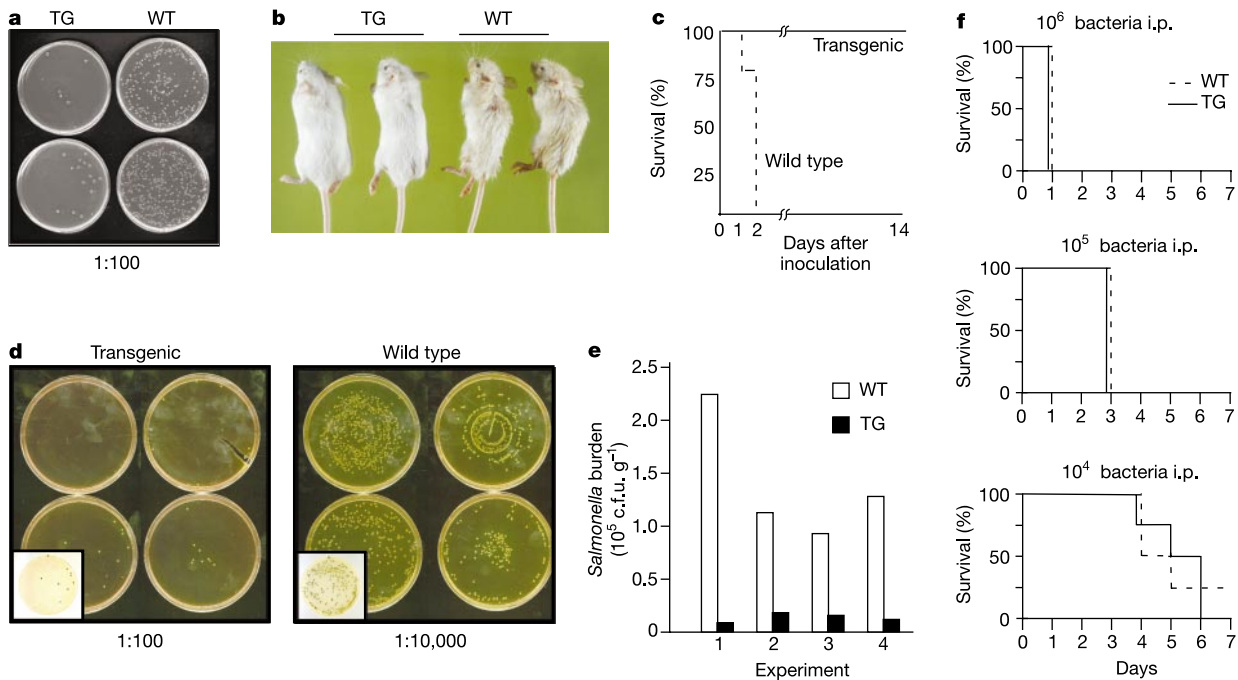


Figure 3 Challenge of HD-5 transgenic and wild-type FvB control mice with virulent *S. typhimurium*. **a**, Comparison of bacterial burden in terminal ileum in HD-5 transgenic and wild-type control mice 6 h after oral challenge with 1×10^8 c.f.u. *S. typhimurium* ($n = 6$ for each group, $P < 0.018$, Student's *t*-test). Representative plates of luminal bacteria recovered and pooled from the terminal ileum of two mice (each plate); bacteria were grown on selective medium. **b**, Comparison of HD-5 transgenic (left) and wild-type mice (right) 12 h after oral challenge with 1.5×10^9 c.f.u. *S. typhimurium*. **c**, Survival curve comparing age-matched HD-5 transgenic with wild-type control mice after oral challenge with 1.5×10^9 c.f.u. *S. typhimurium* ($n = 6$ for each group). **d**, Comparison of

bacterial burden in terminal ileum in HD-5 transgenic and wild-type mice 24 h after oral challenge with 1.5×10^9 c.f.u. *S. typhimurium* (each plate represents bacteria from a single mouse). **e**, Comparison of bacterial burden in spleen of HD-5 transgenic (open) and wild-type control mice (filled), 3 days after oral challenge with 1×10^8 c.f.u. *S. typhimurium*. Four independent experiments ($n = 17$ mice for each group; pooled spleen tissue from 2–4 mice per experiment) are shown (paired *t*-test, $P = 0.026$). **f**, Survival curve comparing HD-5 transgenic (solid line) and wild-type control mice (dotted line) after intraperitoneal (i.p.) challenge with *S. typhimurium* ($n = 4$ for each group).

human intestinal lumen, where intestinal trypsin was identified as the processing enzyme⁹. Proteolytic cleavage at a canonical trypsin site is also observed for the endogenous cryptdin, cryptdin 4, but the identity of the responsible protease for the cleavage of cryptdin 4 and for the transgenic HD-5 remains to be determined. It is probable that HD-5 and cryptdin 4 are processed by the same protease, and future experiments are aimed at identifying this protein. Characterizing the processed form of transgenic HD-5 in mice deficient in MMP-7 may help to clarify this issue.

Several putative roles have been proposed for Paneth cell defensins and other antimicrobial peptides, including protection of the crypt stem cell, the regulation of the numbers and composition of the luminal microbiota, and immediate host defence against food- and water-borne enteric pathogens. In this study, we find that the expression of HD-5 in murine Paneth cells augments the endogenous innate immune capacity of mice to confer marked resistance to oral *Salmonella* infection. This is manifested by reductions in the bacterial burden in the intestinal lumen and faeces, decreased bacterial translocation, and higher survival rates after lethal *Salmonella* challenge. These results clearly support the hypothesis that Paneth cell defensins contribute significantly to mucosal host defence against infectious enteritis. Although this conclusion is more speculative, these results suggest that the natural array of antimicrobial peptides in each animal species may in part determine the pathogenicity of a particular microbe in that species. The HD-5 model may provide a means to test further this and other proposed functions of human enteric defensin peptides.

Our data offer compelling support for the emerging role of mammalian antimicrobial peptides in host defence against bacterial

challenge³. Other reports have described complementary approaches using homologous recombination to eliminate the expression of antimicrobial peptides. Mice rendered deficient in the expression of the antimicrobial peptide cathelin-related antimicrobial peptide (CRAMP) were more susceptible to skin infections by group A *Streptococcus*²³. Furthermore, mice deficient in the metalloprotease matrilysin, an enzyme with several functions including the processing of small intestinal defensins, are highly susceptible to oral infection by *Salmonella*¹³. However, a limitation to the loss-of-function approach for establishing *in vivo* function is that the targeted proteins often have other physiological roles that could indirectly contribute to susceptibility to infection. Here, we report a gain-of-function transgenic model, which provides data on the *in vivo* function of a human defensin peptide. Together, the complementary transgenic approaches reported previously¹³ and reported here establish a key role for mucosal defensins in host defence against orally ingested pathogens. Given that antimicrobial peptides are encoded by conventional genes, future therapies may include the expression of peptides in selected cells and tissues through somatic cell gene transfer. The data reported here complement recent studies on the expression of human cathelicidins and defensins through gene transfer *in vitro*, which yielded increased antimicrobial capabilities of recipient cells^{24–26}. Our data highlight the effectiveness of transferring xenobiotic antimicrobial peptide genes in efforts to augment mucosal host defence. □

Methods

Generation of transgenic mice

Studies reported here were performed using protocols approved by the animal care and utilization committees at the authors' respective institutions. The DNA fragment used for

the transgene was HG2-3e, a 2.9-kb *EcoRI* fragment of the genomic clone, containing the *HD-5* gene and flanking sequences⁷. The genomic fragment was microinjected into fertilized FvB mouse oocytes²⁷ and offspring were produced. DNA was isolated from portions of the tails of the FvB mouse pups and analysed by PCR using HD-5 gene-specific probes HD-5-2172s (CGGCATTTTCAGAACTGATT) and HD-5-2582a (TTCGGCAATAGCAGGTGGCT), using the following conditions in a Perkin Elmer Model 480 Thermocycler: denaturation at 94 °C for 5 min, followed by 35 cycles of 30 s at 94 °C, 30 s at 55 °C, 1 min at 72 °C, generating a 421-bp product. Three mice expressing the transgene were bred against a wild-type FvB mouse, producing independent lines of heterozygous offspring (1661, 6571, 6568) on a pure FvB genetic background. All studies reported here used mice of a pure FvB background. Homozygous mice were generated through crossbreeding of heterozygotes and were identified by genomic Southern blot analysis.

The determination of transgene copy number was made by a method described previously²⁸. We made serial dilutions of the HD-5 genomic fragment. Wild-type mouse genomic DNA was used as a carrier in the dilutions of the HD-5 genomic fragment of DNA. The serial dilutions of the cloned fragment were run on a gel along with 10 µg of liver DNA from each transgenic founder line. Southern blots were hybridized to a HD-5 gene-specific probe Sig68 (GAGTGGCTCAGCCTGGGCTGCAGGGCCACCAGGAGAATGG CAGCAA). The labelled bands were quantified using a phosphorimager. A probe to G3PDH was used to normalize for any variability in loading of DNA.

The genomic sequence encompassing the Paneth cell α -defensin genes *HD-5* (AF228730) and *HD-6* (AF314060) were compared using Pustell analysis with parameters similar to those described^{12,29}. Total RNA isolated from selected mouse tissues were probed with an HD-5 probe (HSIA-309a) as described⁷. Endogenous cryptidin mRNA was detected using a probe complementary to six cryptidins (GTTTATGCTCTT CATCTGTGTTTGGATAGGATCAGCCTGGACCTGG). The probe specific for cryptidin 4 mRNA was GCGGGGGCAGCAGTACAAAATCGTATT CCACAAAGTCCCACGA.

Isolation of HD-5 peptides

Small intestinal tissue from transgenic mice, fasted 12 h before they were killed, was homogenized in 20% acetic acid buffer, neutralized (pH 6) in the presence of a cocktail of protease inhibitors, diluted (1:10) and then applied to a carboxymethylcellulose (CM) resin (BioRad) using methods described previously^{9,17}. To analyse and compare the tissue and luminal distribution of transgenic HD-5 peptide, 4–6-week-old mice were injected with the acetylcholine agonist aceclidine (10 µg g⁻¹ in water, intraperitoneally)³⁰ and killed after 30 min. The small intestinal tissue was bisected lengthwise and the luminal surface was thoroughly flushed with cold normal saline; the resulting wash was diluted with ammonium acetate buffer (pH 6.0) and cationic peptides were adsorbed to CM resin. The cationic components were eluted from the CM resin in 10% acetic acid. An aliquot of the CM eluate was resolved by 12.5% acid urea gel electrophoresis, electro-blotted to a 0.2-µm polyvinylidene fluoride (PVDF) membrane, and analysed using a polyclonal HD-5 antibody and chemiluminescence detection⁹. Cationic components from the CM eluate were further fractionated using a PolyCat A weak cation exchange high-performance liquid chromatography (HPLC) column (PolyLC) as described previously⁹. HD-5 immunoreactive fractions were applied to a C-18 RP-HPLC (Vydac) column eluted with a 45-min linear gradient from 0% to 80% solvent B (0.08% trifluoroacetic acid in acetonitrile). The major HD-5 immunoreactive peak from luminal samples were fractionated by acid urea gel electrophoresis, electro-blotted to a 0.2-µm PVDF membrane, and analysed by N-terminal sequence analysis. Cationic peptides from transgenic and wild-type intestinal lumen, isolated from CM resin, were also assayed using a c.f.u. assay with *E. coli* ML35 as described^{9,10}.

Bacterial challenges

Bacterial inocula of *S. typhimurium* were grown in tripticase soy broth (TSB) to mid-log phase from single colonies. Bacteria were pelleted, re-suspended in fresh TSB (or PBS) and quantified using a Petroff–Hauser counting chamber. *Salmonella typhimurium* 14028s was passaged by oral inoculation in wild-type mice to enhance virulence. HD-5 homozygous FvB mice and age- and sex-matched FvB wild-type control mice were used for challenge experiments. Mice were deprived of food and bedding for 8–14 h before inoculation. Oral inoculation from a disposable syringe, or gentle gavage using a small gavage needle, was used for oral challenges (1–5 × 10⁸ c.f.u. of bacteria in 100 µl TSB or PBS). The mice were then returned to cages with food and bedding. If three inocula were administered, the mice were fed for 1 h after each inoculum, and after the third round of inoculum the mice were returned to cages with food and bedding. Body weight and activity were carefully monitored; moribund mice were killed and analysed by necropsy. For the comparison of *S. typhimurium* burden in the terminal ileum, mice were killed and the terminal ileum was flushed, diluted and plated on *Salmonella*–*Shigella* (SS) agar (Becton Dickinson). Faecal samples were collected fresh from the animals 6–24 h after inoculation, diluted and plated on SS agar. For bacterial translocation studies, the mice were treated as above and 3 days after a single inoculation of 1 × 10⁸ c.f.u., they were killed, their spleens isolated and homogenized, and bacteria were plated in dilution on SS agar. For intraperitoneal infection, HD-5 homozygous mice and wild-type controls (*n* = 4 per group, 6 weeks of age) were injected with 100 µl of PBS containing 10⁶, 10⁵ or 10⁴ c.f.u. of *Salmonella*. The mice were carefully monitored and immediately killed if they became moribund.

Received 9 December 2002; accepted 18 February 2003; doi:10.1038/nature01520.
Published online 19 March 2003.

1. Boman, H. G. Peptide antibiotics and their role in innate immunity. *Annu. Rev. Immunol.* **13**, 61–92 (1995).

- Meister, M., Hetru, C. & Hoffmann, J. A. The antimicrobial host defense of *Drosophila*. *Curr. Top. Microbiol. Immunol.* **248**, 17–36 (2000).
- Zaslöff, M. Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389–395 (2002).
- Lehrer, R. I., Bevins, C. L. & Ganz, T. in *Mucosal Immunology* (eds Ogra, P. L., Mestecky, J., Lamm, M. E., Strober, W. M. & Bienstock, J.) 89–99 (Academic, New York, 1998).
- Ayabe, T. *et al.* Secretion of microbicidal α -defensins by intestinal Paneth cells in response to bacteria. *Nature Immunol.* **1**, 113–118 (2000).
- Porter, E. M., Bevins, C. L., Ghosh, D. & Ganz, T. The multifaceted Paneth cell. *Cell. Mol. Life Sci.* **59**, 156–170 (2002).
- Jones, D. E. & Bevins, C. L. Paneth cells of the human small intestine express an antimicrobial peptide gene. *J. Biol. Chem.* **267**, 23216–23225 (1992).
- Porter, E., Liu, L., Oren, A., Anton, P. & Ganz, T. Localization of human intestinal defensin 5 in Paneth cell granules. *Infect. Immun.* **65**, 2389–2395 (1997).
- Ghosh, D. *et al.* Paneth cell trypsin is the processing enzyme for human defensin-5. *Nature Immunol.* **3**, 583–590 (2002).
- Porter, E., van Dam, E., Valore, E. & Ganz, T. Broad spectrum antimicrobial activity of human intestinal defensin 5. *Infect. Immun.* **65**, 2396–2401 (1997).
- Ouellette, A. J. & Bevins, C. L. in *Development of the Gastrointestinal Tract* (eds Sanderson, I. R. & Walker, W. A.) 147–164 (Decker, Hamilton, Ontario, 1999).
- Mallow, E. B. *et al.* Human enteric defensins: gene structure and developmental expression. *J. Biol. Chem.* **271**, 4038–4045 (1996).
- Wilson, C. L. *et al.* Regulation of intestinal α -defensin activation by the metalloproteinase matrilysin in innate host defense. *Science* **286**, 113–117 (1999).
- Darmoul, D. & Ouellette, A. J. Positional specificity of defensin gene expression reveals Paneth cell heterogeneity in mouse small intestine. *Am. J. Physiol.* **271**, G68–G74 (1996).
- Wang, M. S., Pang, J. S. & Selsted, M. E. Semidry electroblotting of peptides and proteins from acid-urea polyacrylamide gels. *Anal. Biochem.* **253**, 225–230 (1997).
- Cole, A. M. & Ganz, T. Human antimicrobial peptides: analysis and application. *Biotechniques* **29**, 822–831 (2000).
- Porter, E. M. *et al.* Isolation of human intestinal defensins from ileal neobladder urine. *FEBS Lett.* **434**, 272–276 (1998).
- Cunliffe, R. N. *et al.* Human defensin 5 is stored in precursor form in normal Paneth cells and is expressed by some villous epithelial cells and by metaplastic Paneth cells in the colon in inflammatory bowel disease. *Gut* **48**, 176–185 (2001).
- Satoh, Y., Habara, Y., Ono, K. & Kanno, T. Carbamylcholine- and catecholamine-induced intracellular calcium dynamics of epithelial cells in mouse ileal crypts. *Gastroenterology* **108**, 1345–1356 (1995).
- Ouellette, A. J., Satchell, D. P., Hsieh, M. M., Hagen, S. J. & Selsted, M. E. Characterization of luminal Paneth cell α -defensins in mouse small intestine: attenuated antimicrobial activities of peptides with truncated amino termini. *J. Biol. Chem.* **275**, 33969–33973 (2000).
- Selsted, M. E., Miller, S. I., Henschen, A. H. & Ouellette, A. J. Enteric defensins: antibiotic peptide components of intestine host defense. *J. Cell Biol.* **118**, 929–936 (1992).
- Putsep, K. *et al.* Germ-free and colonized mice generate the same products from enteric prodefensins. *J. Biol. Chem.* **275**, 40478–40482 (2000).
- Nizet, V. *et al.* Innate antimicrobial peptide protects the skin from invasive bacterial infection. *Nature* **414**, 454–457 (2001).
- Couto, M. A., Liu, L., Lehrer, R. I. & Ganz, T. Inhibition of intracellular *Histoplasma capsulatum* replication by murine macrophages that produce human defensin. *Infect. Immun.* **62**, 2375–2378 (1994).
- Bals, R., Weiner, D. J., Moscioni, A. D., Meegalla, R. L. & Wilson, J. M. Augmentation of innate host defense by expression of a cathelicidin antimicrobial peptide. *Infect. Immun.* **67**, 6084–6089 (1999).
- Kisich, K. O., Heifets, L., Higgins, M. & Diamond, G. Antimycobacterial agent based on mRNA encoding human β -defensin 2 enables primary macrophages to restrict growth of *Mycobacterium tuberculosis*. *Infect. Immun.* **69**, 2692–2699 (2001).
- Taketo, M. *et al.* FVB/N: an inbred mouse strain preferable for transgenic analyses. *Proc. Natl Acad. Sci. USA* **88**, 2065–2069 (1991).
- Harper, M. E., Barrera-Saldana, H. A. & Saunders, G. F. Chromosomal localization of the human placental lactogen-growth hormone gene cluster to 17q22-24. *Am. J. Hum. Genet.* **34**, 227–234 (1982).
- Bevins, C. L., Jones, D. E., Dutra, A., Schaffzin, J. & Muenke, M. M. Human enteric defensin genes: chromosomal map position and a model of possible evolutionary relationships. *Genomics* **31**, 95–106 (1996).
- Iwabuchi, Y., Katagiri, M. & Masuhara, T. Salivary secretion and histopathological effects after single administration of the muscarinic agonist SNI-2011 in MRL/lpr mice. *Arch. Int. Pharmacodyn. Ther.* **328**, 315–325 (1994).

Acknowledgements We acknowledge the technical assistance of J. Russell, S. Lee, D. Wilk, H. deJong and K. Carver. We thank J. Crabb and S. Yadav for their help with peptide analysis, and T. Ganz and E. Porter for discussions and for providing us with recombinant peptides and the polyclonal anti-HD-5 antibody. We also acknowledge discussions with S. Miller and A. Ouellette. This work was supported by grants from the National Institutes of Health (to C.L.B. and N.H.S.) and The Crohn's and Colitis Foundation of America (to N.H.S.). A portion of this work was completed while the indicated authors were at the Departments of Microbiology (N.H.S.) and Pediatrics (N.H.S., C.L.B.), University of Pennsylvania, School of Medicine and the Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.L.B. (e-mail: bevins@ccf.org).

PAC1 phosphatase is a transcription target of p53 in signalling apoptosis and growth suppression

Yuxin Yin*, Yu-Xin Liu*, Yan J. Jin*, Eric J. Hall* & J. Carl Barrett†

* Department of Radiation Oncology, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, New York 10032, USA

† Laboratory of Biosystems and Cancer, Center for Cancer Research, National Cancer Institute, NIH, 31 Center Drive, Bethesda, Maryland 20892, USA

p53 has a role in many cellular processes through the transcriptional regulation of target genes^{1,2}. PAC1 (phosphatase of activated cells 1; also known as dual specificity phosphatase 2, DUSP2) is a dual threonine/tyrosine phosphatase that specifically dephosphorylates and inactivates mitogen-activated protein (MAP) kinases^{3,4}. Here we show that during apoptosis, p53 activates transcription of PAC1 by binding to a palindromic site in the PAC1 promoter. PAC1 transcription is induced in response to serum deprivation and oxidative stress, which results in p53-dependent apoptosis, but not in response to γ -irradiation, which causes cell cycle arrest^{5,6}. Reduction of PAC1 transcription using small interfering RNA inhibits p53-mediated apoptosis, whereas overexpression of PAC1 increases susceptibility to apoptosis and suppresses tumour formation. Moreover, activation of p53 significantly inhibits MAP kinase activity. We conclude that, under specific stress conditions, p53 regulates transcription of PAC1 through a new p53-binding site, and that PAC1 is necessary

and sufficient for p53-mediated apoptosis. Identification of a palindromic motif as a p53-binding site may reveal a novel mechanism whereby p53 regulates its target genes.

Using the DNA microarray technique, we previously identified more than 60 genes upregulated by p53 (ref. 7). Human PAC1, which inactivates extracellular signal-regulated kinase 1 and 2 (ERK1/2) activity through dephosphorylation of ERK1/2 (ref. 4) is among those p53 target genes. Therefore, it is important to dissect the molecular link between the tumour suppressor gene p53 and an essential cell signalling pathway: the MAP kinase cascade⁸. To confirm the data from the microarray study, we performed northern analysis of PAC1 expression in a p53-inducible system using EB-1 cells. The EB cell line was derived from a human colon cancer with mutant p53. EB-1 is a stable clone containing a wild-type p53 transgene under the control of the metallothionein promoter and expressing wild-type p53 on administration of zinc chloride (ZnCl₂)⁹. These cells undergo apoptosis after serum deprivation in the presence of p53 (ref. 9). As expected, p21, a p53 target gene, is absent in EB-1 without wild-type p53 (Fig. 1a, middle panel, lane 1) but increased after wild-type p53 expression induced by ZnCl₂ (Fig. 1a, middle panel, compare lanes 1 and 2). PAC1 transcript levels are low in EB-1 cells without p53 (Fig. 1a, top panel, lane 1). Notably, PAC1 expression is not induced in EB-1 cells with administration of ZnCl₂ (compare lanes 1 and 2), or through γ -irradiation, which induces the expression of p21 (Fig. 1a, middle panel, lanes 2 and 5) and inhibits cell cycle progression (Fig. 1e). However, PAC1 expression is greatly increased in EB-1 cells treated with both ZnCl₂ and serum deprivation (Fig. 1a, top panel, lane 3), which activates p53 and induces apoptosis (see Fig. 1c)⁹. In addition, PAC1 expression is also induced by H₂O₂ in the presence

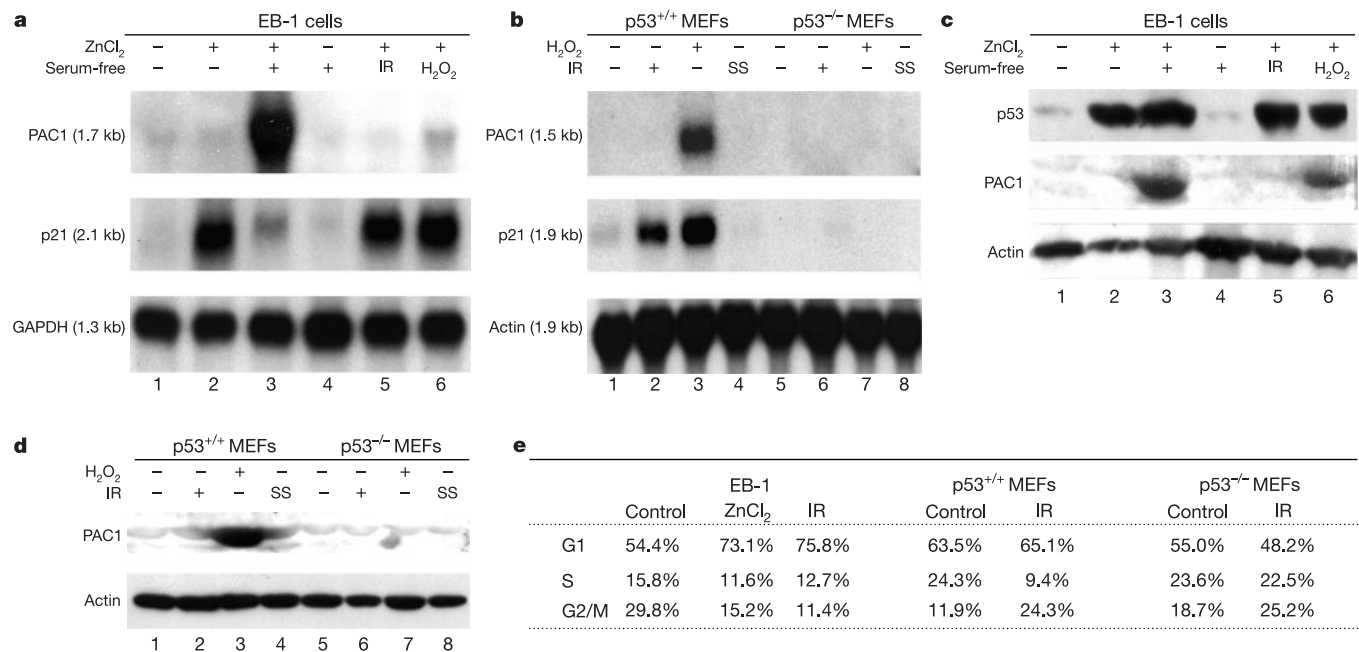


Figure 1 Transcriptional regulation of PAC1 by p53. **a**, Induction of human PAC1 expression during p53-mediated apoptosis by serum deprivation and H₂O₂. EB-1 cells were cultured under the following conditions: MEM medium with 10% FBS (lane 1); MEM with 10% FBS plus 100 μ M ZnCl₂ (lane 2); MEM with 0.1% FBS plus 100 μ M ZnCl₂ (lane 3); MEM with 0.1% FBS without ZnCl₂ (lane 4); MEM with 10% FBS plus 100 μ M ZnCl₂ and 6 Gy of γ -irradiation (IR, lane 5) or 200 μ M H₂O₂ (lane 6). **b**, Expression of PAC1 in MEFs under various conditions. Exponentially growing MEFs were exposed to 6 Gy of γ -irradiation, 100 μ M H₂O₂, or serum starvation (SS), as indicated. The cells from **a** and **b** were collected after 3 h, and mRNA was fractionated on a 1.2% formaldehyde agarose gel

and transferred for northern blotting using a [α -³²P]dCTP-labelled PAC1 or p21 cDNA probe, respectively, following standard procedures. **c**, **d**, Induction of PAC1 protein by p53 under stress conditions. Exponentially growing EB-1 cells (**c**) and MEFs (**d**) were treated as indicated. Cell extracts were resolved by 10% SDS-PAGE and transferred for western blotting using an anti-p53 antibody (PAb421) or anti-PAC1 antibodies (Santa Cruz). **e**, p53-mediated cell cycle response to γ -irradiation. EB-1 cells and MEFs were exposed to γ -irradiation (6 Gy) and collected after 16 h. The cells were incubated with propidium iodide and processed for flow cytometric analysis. Distribution of cells in the G1, S and G2/M phases is represented as percentages.

of p53 (Fig. 1a, lane 6), which causes oxidative cell death (data not shown). Correspondingly, PAC1 protein levels are increased by p53 on serum starvation and oxidative stress (see Fig. 1c, lanes 3, 6). These results suggest that PAC1 transcription is regulated by p53, and that induction of PAC1 is dependant on p53 status and influenced by stress conditions.

To determine whether p53 is required for the PAC1 response to stress conditions, we treated normal and p53 null mouse embryo fibroblasts (p53^{+/+} and p53^{-/-} MEFs, respectively) with γ -irradiation, which causes cell cycle arrest at G1 (ref. 5) or G2 (Fig. 1e), or with H₂O₂, which induces cell death by apoptosis¹⁰.



Figure 2 Transactivation of the PAC1 promoter by p53. **a**, Sequence of the human PAC1 promoter. Nucleotide sequence upstream of the translation start site was annotated from a PCR-amplified regulatory region of human PAC1. The approximate transcription start site is shown as an arrow. A palindromic site is indicated by sequence that is bold and underlined. Numbering is with respect to the translation initiation site. **b**, Induction of promoter activities of PAC1 and p21 by wild-type p53. H1299 cells were transiently transfected with respective plasmids described below using LipofectAMINE in Opti-MEM medium (Gibco). Cell extracts were assayed for luciferase activity on a Berthold autolumat LB953 rack luminometer. The luciferase activity readout is expressed as means $\pm$ s.d. of triplicate cultures and transfections. The transfection groups (from left to right) are as follows: empty pCMV vector; pCMV-wtp53 plus pGL3-basic; pCMV vector plus pGL3-PAC1-0.7; pCMV-wtp53 plus pGL3-PAC1-0.7; pCMV-248 plus pGL3-PAC1-0.7; pCMV plus p21 promoter luciferase reporter pGL3-p21; pCMV-wtp53 plus pGL3-p21; pCMV-p53Δ370 plus pGL3-PAC1-0.7; pCMV-wtp53 plus a mutated PAC1 reporter pGL3-PAC1-0.7m.

PAC1 expression is low in p53^{+/+} MEFs and is not induced by γ -irradiation (Fig. 1b, top panel, lanes 1, 2). PAC1 is markedly increased after oxidative damage by H₂O₂ (Fig. 1b, top panel, compare lanes 1 and 3). However, PAC1 expression is not induced by serum starvation (Fig. 1b, lane 4), which does not induce cell death in MEFs (data not shown). Furthermore, the response of PAC1 to oxidative damage is completely diminished in p53^{-/-} MEFs (lane 3 and lane 7). Similarly, the level of PAC1 protein is

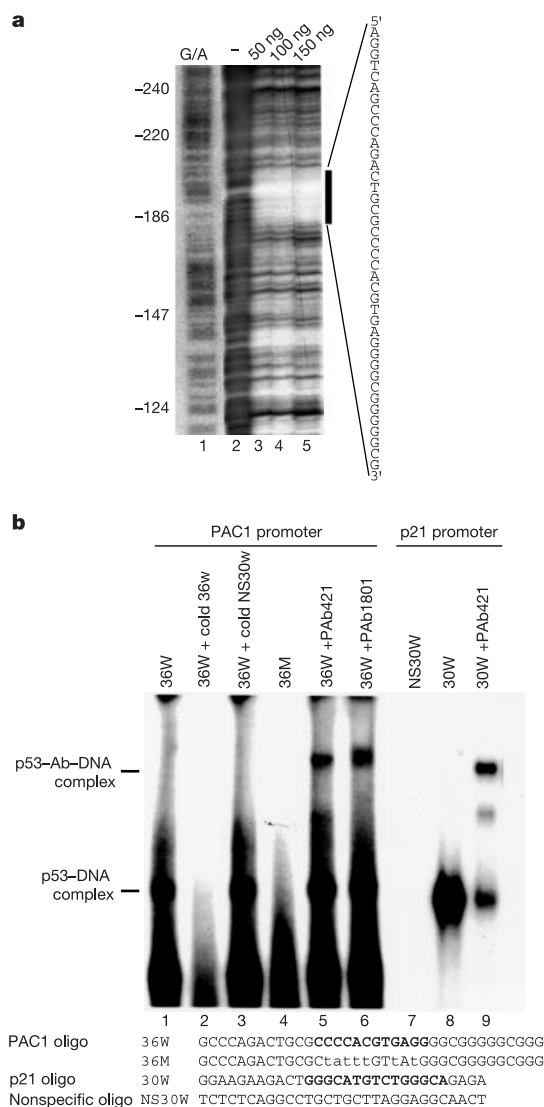


Figure 3 Physical interaction between p53 and a palindromic site in the PAC1 promoter. **a**, Footprint analysis of the p53-binding site in the PAC1 promoter. The end-labelled promoter fragment was incubated with DNase I in the absence or presence of baculovirus-expressed wild-type p53 and the reactions were electrophoresed in sequencing gels as described in Methods. The G/A ladders were prepared by using the Maxam-Gilbert sequencing method as footprint markers. The DNase I footprint between -170 and -210 is indicated as an empty region. The corresponding sequencing within the footprint is listed as one strand from 5' to 3'. **b**, Gel shift analysis of the interaction between p53 protein and oligonucleotides containing a palindromic site. Purified recombinant human wild-type p53 protein was incubated with ³²P-labelled double-stranded oligonucleotide containing the palindromic sequence. For competition, 50-fold excess of the unlabelled oligonucleotide was included in reactions. Supershift was carried out by adding anti-p53 antibodies (PAb421 or PAb1801) in reactions. The protein-DNA complexes were separated by 4% native PAGE and visualized by autoradiography. All oligonucleotide sequences are listed as one strand (5' to 3').

greatly increased in p53^{+/+} but not in p53^{-/-} MEFs after oxidative stress (Fig. 1d, compare lanes 3 and 7). By comparison, p21 is significantly induced by p53 after treatment of either γ -irradiation or oxidative stress (Fig. 1b, middle panel, lanes 2, 3). These data suggest that PAC1 expression is inducible by oxidative stress in a p53-dependent manner and that p53 can selectively regulate PAC1 in response to oxidative damage.

To understand the molecular basis for regulation of PAC1 by p53, we cloned a regulatory region of the human *PAC1* gene. As partially listed in Fig. 2a, sequence analysis of the promoter region reveals no known p53 consensus binding site as 5'-PuPuPuC(A/T)(T/A)GPpPyPy-3'¹¹. However, there is a 12-base pair (bp) palindromic site from -202 to -191 (5'-CCCCACGTGAGG-3') in the PAC1 promoter, as indicated by underlined letters (Fig. 2a).

To determine whether the PAC1 promoter is regulated by p53, the PAC1 promoter was amplified from -726 to -15, which contains the palindromic sequence. This 711-bp promoter was ligated into a luciferase reporter vector pGL3-basic (Promega) adjacent to a luciferase reporter gene, resulting in the pGL3-PAC1-0.7 reporter. To examine luciferase activity, the PAC1 luciferase reporter and other reporters were transfected into H1299 cells along with a pCMV vector as a control, or with a wild-type p53 expression vector (pCMV-wtp53)¹². As shown in Fig. 2b, there is a basal activity of the PAC1 promoter in the absence of p53. The activity of pGL3-PAC1-0.7 is markedly induced by wild-type p53, but not by a mutant p53 plasmid (pC53-248) that contains a point mutation at codon 248, suggesting that the PAC1 promoter is highly responsive to p53 function. A lysine-rich basic domain at the carboxy terminus of p53 inhibits specific DNA binding properties of p53 (ref. 13). We therefore tested whether this inhibitory domain influences p53-induced activation of the PAC1 promoter by using an expression vector with truncated p53 lacking the C-terminal lysine-rich basic domain (pCMV-p53 Δ 370)¹⁴ in a luciferase assay. We observed that expression of the PAC1 promoter is increased with this truncated p53 to a level similar to that of wild-type p53, suggesting that the C-terminal domain of p53 does not interfere with the binding of p53 to the PAC1 promoter. To test whether the palindromic sequence is important to p53 transactivity, we constructed a series of luciferase reporters with or without the palindromic region. We observed that luciferase activity was reduced greatly in all of the luciferase reporters omitting the palindromic sequence in the presence of p53 (data not shown). These results suggest that p53 regulates the PAC1 promoter activity probably through a region containing this palindromic site. To test further this possibility, we performed a polymerase chain reaction (PCR)-based site-directed mutagenesis to create a mutant form of pGL3-PAC1-0.7 (pGL3-PAC1-0.7m), which lacks a 6-bp core sequence (CCCCAC) of the palindrome. The p53-induced luciferase activity of this mutant reporter was reduced to the basal level (Fig. 2b). These results demonstrate that the palindromic site is crucial for induction of the PAC1 promoter activity by p53.

To identify the minimum p53 binding sequence, we conducted a DNase I footprinting analysis to pinpoint the binding site of the

PAC1 promoter using a Core Footprinting System (Promega). As shown in Fig. 3a, there is a clear and well-defined footprint region between -170 to -210 of the PAC1 promoter in the presence of p53 (compare lane 2 with 3–5), which includes the palindromic site. To determine further whether p53 physically binds to this palindromic site, we synthesized a 36-bp oligonucleotide containing the palindromic sequence (36W, see oligonucleotide list in Fig. 3b, bottom panel) and used it as a probe to perform an electrophoretic mobility shift assay (EMSA). As shown in Fig. 3b, one prominently shifted band was detected when this 36-bp radiolabelled element was incubated with recombinant human wild-type p53 (lane 1), which was competed out by excess unlabelled 36W oligonucleotide (lane 2). There was no band shift seen when p53 was incubated with 36M, a mutated form of the 36W oligonucleotide containing an incomplete sequence of the palindromic site (lane 4), nor a non-specific oligonucleotide (lane 7). To confirm p53-binding specificity, two anti-p53 monoclonal antibodies, PAB421 and PAB1801, were included in the reactions, respectively. The addition of either of the two p53 antibodies resulted in a clear 'supershifted' band (lanes 5, 6). As a positive control, we included a 30-bp synthetic oligonucleotide containing a consensus p53-binding motif from the p21 promoter¹⁵ in the EMSA reactions. As expected, p53 formed a shift band with this oligonucleotide (lane 8), which was further shifted in the presence of the PAB421 antibody (lane 9). These results suggest that p53 specifically binds to the oligonucleotide containing the palindromic site, as p53 binds its consensus site in the p21 promoter.

To demonstrate the importance of PAC1 in signalling cell death, we used RNA interference technology¹⁶ to reduce PAC1 expression. Using a mammalian expression vector for small interfering RNA (siRNA), pSilencer 1.0-U6 siRNA expression vector¹⁷, we constructed human and mouse PAC1 siRNA vectors and introduced these into either EB-1 cells or MEFs. The stable clones were tested for PAC1 expression by northern blotting (Fig. 4a). In contrast to wild-type cells, those transfected with PAC1 siRNA contain undetectable PAC1 transcripts under conditions of oxidative stress (Fig. 4a, compare lane 2 with 3–6). As shown in Fig. 4b, the MEFs transfected with PAC1 siRNA, in which PAC1 messenger RNA levels were undetectable, were highly resistant to cell killing by H₂O₂ treatment. Furthermore, in contrast to the EB-1-U6 control cells (Fig. 4h, i), EB-1 cells transfected with PAC1 siRNA-c5, which display a low level of PAC1 expression when p53 is induced by ZnCl₂ treatment (data not shown), failed to undergo apoptosis on p53 activation by serum deprivation (Fig. 4j, k). These results demonstrate that PAC1 has an essential role in the p53-mediated apoptotic response to oxidative damage and nutritional stress.

To determine the role of PAC1 in the apoptotic process, a human PAC1 expression vector, pcDNA3-PAC1, was transfected into EB cells and a breast cancer cell line, MDA-MB435, which were then selected for isolating stable clones of EB-PAC1 (Fig. 4c, lanes 2, 3) and MB435-PAC1 (Fig. 4c, lanes 5–7). We found that EB-PAC1-c5 cells, which contain a high level of ectopic PAC1 (Fig. 4c, bottom panel, lane 2), undergo typical apoptosis under the condition of serum deprivation (Fig. 4d–g), and that MB435-PAC1 cells, par-

Table 1 Inhibition of tumorigenicity of MDA-MB435 cells by PAC1

Cancer cell lines	No. colonies in soft agar ($\pm$ s.d.)	No. of mice with tumours			Terminal mean tumour volume (cm ³) ($\pm$ s.d.)
		1 week	2 weeks	4 weeks	
MB435 (parental)	600 $\pm$ 20	8/10	10/10	10/10	2.6 $\pm$ 0.5
MB435-pcDNA3	580 $\pm$ 30	6/10	10/10	10/10	2.9 $\pm$ 0.9
MB435-PAC1-C6	230 $\pm$ 30	1/10	3/10	3/10	0.9 $\pm$ 0.2
MB435-PAC1-C7	360 $\pm$ 15	0/10	3/10	4/10	1.2 $\pm$ 0.2
MB435-PAC1-C12	373 $\pm$ 20	0/10	1/10	4/10	0.6 $\pm$ 0.2

The cells from each group were first tested for anchorage-independent colony formation in soft agarose as describe in Methods. Colonies with a diameter of >1.0 mm were counted after 2 weeks of culture. For each inoculation, 5 $\times$ 10⁵ cells were injected subcutaneously on the back region of athymic nude mice (BALB/c/nu/nu). The number of tumours was counted after 4 weeks and size of tumours was determined according to the equation $V = (L \times W^2) \times 0.5$, where L represents length and W represents width of a tumour. Statistical significance between MB435 control groups and PAC1 clones was evaluated by Student's t -test ($P < 0.05$).

ticularly, MB435-PAC1-c6, are highly susceptible to cell death as a result of H_2O_2 treatment (Fig. 4l). These results indicate that PAC1 functions as a cell death mediator in the cellular response to oxidative damage or nutritional stress. To test the possibility that p53 influences the activity of MAP kinases, we examined the phosphorylation status of myelin basic protein (MBP)—a specific substrate for MAP kinases—in the presence of immunoprecipitated ERK1/2 from EB-1 cells and MB435-PAC1 clones. As shown in Fig. 4m, MAP kinase activity was decreased after activation of p53 (lane 2). As expected, MAP kinase activity was decreased in MB435-PAC1 cells (lanes 5, 6). These results suggest that p53 is capable of suppressing the MAP kinase cascade and signalling down a diverse cell survival pathway, probably through transcriptional regulation of PAC1.

To determine whether overexpression of PAC1 influences anchorage-independent growth, we compared MB435 cells and MB435 cells overexpressing PAC1 for their ability to form colonies in soft agar. As shown in Table 1, the colony numbers from MB435-PAC1 cells in soft agar are significantly less than that from the MB435-pcDNA3 control group. Because overexpression of PAC1 induces apoptosis and inhibits anchorage-independent growth of MB435 cells, it is possible that PAC1 influences tumorigenesis *in vivo*. We tested MB435 cells with or without ectopic PAC1 for their ability to form tumours subcutaneously in athymic nude mice. As shown in Table 1, all mice inoculated with MB435 or MB435-pcDNA3 cells

produced tumours. In contrast, only 30–40% of mice inoculated with MB435-PAC1 clones formed tumours, and their size was much smaller than that of the controls. Our results support the hypothesis that p53 suppresses tumorigenesis mainly through induction of cell death in tumours, and provide evidence that PAC1 is a critical downstream effector of p53 action in initiating apoptosis and tumour suppression.

We show here that PAC1 is a transcriptional target of p53 in signalling apoptosis and tumour suppression *in vivo*. PAC1 is the first reported p53 downstream gene that is required for p53-mediated apoptosis in response to oxidative stress. Because p53 is a primary activator of PAC1 and PAC1 is a principal element that negatively regulates MAP kinase activity, our data establish a molecular link between the p53 tumour suppressor and the MAPK signal transduction cascade. Thus, it is possible that p53 can control cell growth signalling through regulation of PAC1 and inhibition of the MAPK cascade, a cell survival pathway. There has been difficulty in establishing how p53 determines cell fate after different types of damage¹. Because PAC1 is selectively regulated by p53 and required for apoptosis, PAC1 is a specific and critical mediator of cell death in the p53 pathway. Our results also demonstrate a new mechanism for transcriptional regulation of PAC1 by p53. It is well known that p53 binds to its 'consensus site' for transcriptional regulation of its target genes^{2,11}; however, this consensus site has been found only in a limited number of the known p53-responsive genes. We demon-

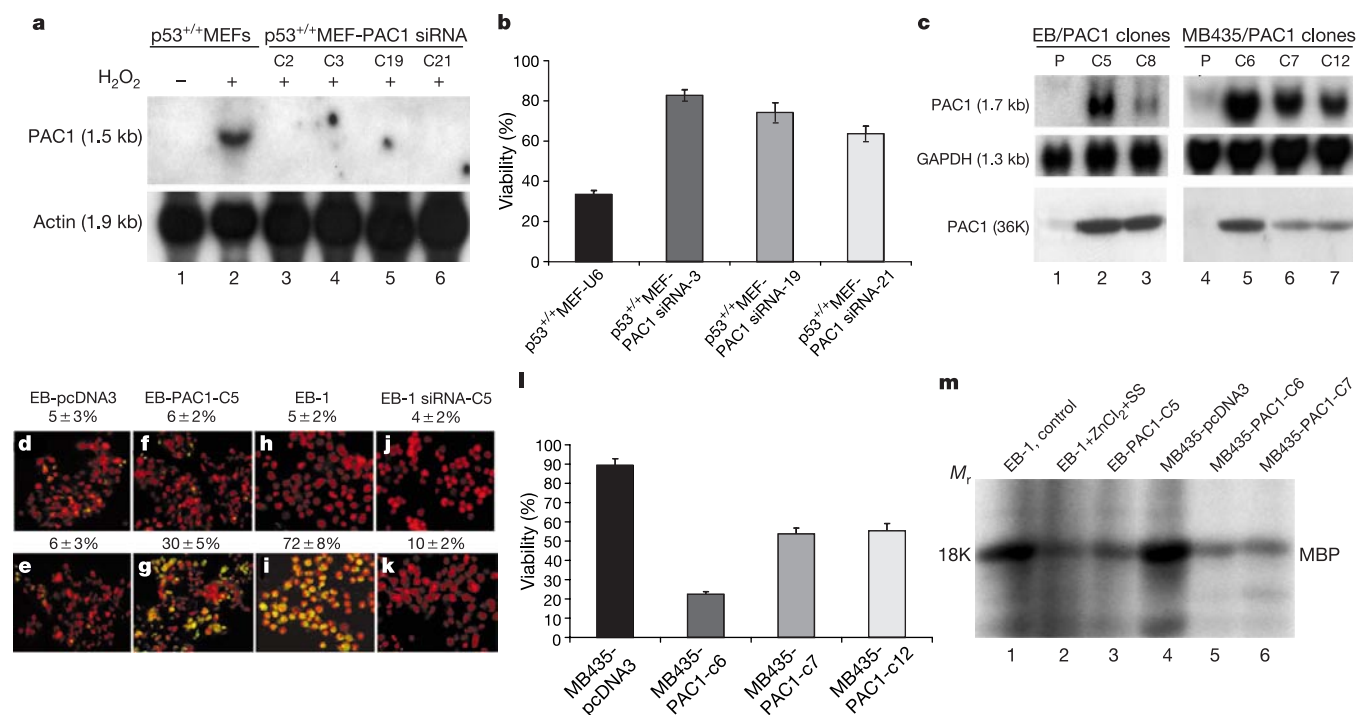


Figure 4 Requirement of PAC1 for p53-mediated apoptosis. **a**, Reduction of PAC1 transcription by expressing PAC1 siRNA in MEFs. MEFs were transfected with either pSilencer U6 or U6-PAC1 siRNA vectors, and stable clones were designated as MEF-U6 or MEF-PAC1 siRNA. The cells were treated with 100 μ M H_2O_2 for 4 h for Northern blotting. **b**, Cell viability of MEFs with or without PAC1 siRNA under oxidative stress. Exponentially growing MEF-U6 or MEF-PAC1 siRNA cells were treated with H_2O_2 (100 μ M) for 24 h. The populations were scored for viable and nonviable cells by Trypan blue exclusion. The results are shown as the means $\pm$ s.d. of three independent experiments each containing duplicate cultures. **c**, Ectopic expression of human PAC1 in EB cells and in MB435 cells. EB and MB435 cells were transfected with either pcDNA3 or pcDNA3-PAC1 and selected for stable clones. Northern blotting was carried out using [α -³²P]dCTP-labelled cDNA probes. Western analysis of PAC1 protein expression was performed using the anti-PAC1 antibody (bottom panel). **d–k**, TUNEL analysis of PAC1-mediated

apoptosis. EB and EB-PAC1-c5 cells were cultured with 10% FBS (**d, f**) or with 0.1% FBS (**e, g**) for 48 h. EB-1 and EB-1-PAC1 siRNA (clone 5) cells were cultured under the following conditions: MEM with 10% FBS plus 100 μ M ZnCl₂ (**h, j**); MEM with 0.1% FBS plus 100 μ M ZnCl₂ (**i, k**) for 6 h. The cells were stained by TUNEL assay using an *in situ* apoptosis detection kit (Intergen). The double-stained cells were counted as apoptotic and are shown as a percentage with s.d. of two independent experiments. **l**, Cell viability of MDA-MB435 cells with or without ectopic PAC1 under oxidative stress. The indicated MDA-MB435 clones were treated with H_2O_2 (200 μ M) for 24 h and were scored for viable cells by Trypan blue exclusion. **m**, Inhibition of MAP kinase activity by p53 and PAC1. Cell lysates were immunoprecipitated by anti-ERK1/2 antibodies and the immunoprecipitated complex was incubated with MBP in the presence of [γ -³²P]ATP as described elsewhere²¹. Phosphorylated MBP was resolved by 15% SDS-PAGE and visualized by autoradiography. M_r , relative molecular mass.

strate here that p53 uses a palindromic binding site to regulate its target gene *PAC1*. Thus, it is conceivable that p53 may selectively regulate different groups of target genes through this mechanism or through conventional mechanisms. The identification of this mechanism for p53 action will provide insights into the molecular basis of how p53 selectively regulates its target genes to eliminate cancer cells and suppress tumorigenesis. □

Methods

Cell culture and DNA transfection

EB cells, EB-1 cells and MEFs have been described elsewhere^{7,18,19}. All cancer cell lines were from American type culture collection (ATCC). The conditions for cell culture are described in Supplementary Information. LipofectAMINE reagent (Gibco) was used for transient and stable transfection of cells, according to the manufacturer's protocol. For selecting stable clones, transfected cells were grown in medium containing $400 \mu\text{g ml}^{-1}$ G418 for neomycin resistance, or $2 \mu\text{g ml}^{-1}$ hygromycin B for hygromycin resistance.

Northern, western and cell cycle analyses

Total RNA was isolated from growing cells using TRIzol reagent (Gibco). Poly(A)⁺ RNA was purified using a olyATtract mRNA isolation system (Promega), according to the manufacturer's instructions. The experimental procedures for northern, western and cell cycle analyses are described in Supplementary Information.

Luciferase reporters, PAC1 siRNA and PAC1 expression plasmids

The regulatory and promoter region of the human *PAC1* gene was cloned by PCR from normal human genomic DNA using corresponding primers based on the human genome database. PCR products were subcloned into a pT-Adv vector (Clontech) for sequencing and compared with the genome database. The PAC1 promoter was subcloned into the pGL3-basic reporter (Promega). The deletion of the palindrome in PAC1 of the pGL3-PAC1-711 plasmid was generated using the QuickChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's protocol. The designs for the constructs, luciferase assay, cellular viability and TdT-mediated dUTP nick end labelling (TUNEL) assay are described in Supplementary Information.

DNase I footprinting analysis

The human PAC1 promoter was end-labelled with [γ -³²P]ATP by T4 polynucleotide kinase, digested with *EcoRV* and purified to obtain the sense strand 3'-end-labelled probe, and subjected to DNaseI footprinting analysis with the Core Footprinting System (Promega), according to the manufacturer's instructions. Recombinant p53 protein was produced in insect cells infected with a baculovirus vector expressing human wild-type p53, and partially purified through affinity chromatography. The purified recombinant p53 protein was added to bind to radiolabelled probe fragment ($1-2 \times 10^4$ c.p.m.) at 37 °C for 30 min, followed by the addition of DNase I. Samples were subjected to polyacrylamide gel electrophoresis (PAGE) under denaturing conditions and the dried gel was exposed to autoradiography with an intensifying screen.

Electrophoretic mobility shift assay

Synthetic oligonucleotides (pairs of sense and antisense) were annealed and labelled with ³²P by using T4 polynucleotide kinase and [γ -³²P]ATP, as described elsewhere²⁰. Briefly, ³²P-labelled probes were mixed with purified recombinant p53 in a 20- μl DNA binding reaction buffer. For specificity or competition controls, a labelled random oligonucleotide or excess of unlabelled corresponding oligonucleotide were added together in reactions. For supershift, the anti-p53 monoclonal antibodies (PAb421, PAb1801; Oncogene) were included. The reaction mixtures were incubated at 4 °C for 20 min, resolved by a 4% polyacrylamide gel, and exposed for photography.

Anchorage-independent growth and tumorigenicity assays

Exponentially growing cells (6×10^3 cells per group) were mixed with 3 ml top agarose containing 0.35% low melting point agarose in MEM medium with 10% fetal bovine serum (FBS), and seeded onto 3 ml 0.6% solidified agarose in the same medium in six-well plates. Colonies with a diameter of >1.0 mm were counted after two weeks of incubation. Results are expressed as means of colony number $\pm$ s.d. of triplicate repeats. For tumorigenicity assay *in vivo*, 4-week-old athymic nude mice (BALB/c/nu/nu, Harlan–Sprague–Dawley) were inoculated subcutaneously with 5×10^5 cells in 0.2 ml MEM medium. Tumour volume was determined by the equation $V = (L \times W^2) \times 0.5$, where L is length and W is width of tumour. Values are the means $\pm$ s.d. of the counted tumours.

Received 17 December 2002; accepted 20 February 2003; doi:10.1038/nature01519.

- Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
- Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
- Rohan, P. J. *et al.* PAC-1: a mitogen-induced nuclear protein tyrosine phosphatase. *Science* **259**, 1763–1766 (1993).
- Ward, Y. *et al.* Control of MAP kinase activation by the mitogen-induced threonine/tyrosine phosphatase PAC1. *Nature* **367**, 651–654 (1994).
- Kastan, M. B. *et al.* A mammalian cell cycle checkpoint pathway utilizing p53 and GADD45 is defective in ataxia-telangiectasia. *Cell* **71**, 587–597 (1992).
- Brugarolas, J. *et al.* Radiation-induced cell cycle arrest compromised by p21 deficiency. *Nature* **377**, 552–557 (1995).
- Zhao, R. *et al.* Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes Dev.* **14**, 981–993 (2000).

- Nishida, E. & Gotoh, Y. The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends Biochem. Sci.* **18**, 128–131 (1993).
- Shaw, P. *et al.* Induction of apoptosis by wild-type p53 in a human colon tumour-derived cell line. *Proc. Natl Acad. Sci. USA* **89**, 4495–4499 (1992).
- Yin, Y., Solomon, G., Deng, C. & Barrett, J. C. Differential regulation of p21 by p53 and Rb in cellular response to oxidative stress. *Mol. Carcinog.* **24**, 15–24 (1999).
- El Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W. & Vogelstein, B. Definition of a consensus binding site for p53. *Nature Genet.* **1**, 45–49 (1992).
- Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. & Vogelstein, B. Suppression of human colorectal carcinoma cell growth by wild-type p53. *Science* **249**, 912–915 (1990).
- Foord, O. S., Bhattacharya, P., Reich, Z. & Rotter, V. A DNA binding domain is contained in the C-terminus of wild type p53 protein. *Nucleic Acids Res.* **19**, 5191–5198 (1991).
- Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90**, 595–606 (1997).
- El-Deiry, W. S. *et al.* Topological control of p21WAF1/CIP1 expression in normal and neoplastic tissues. *Cancer Res.* **55**, 2910–2919 (1995).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Suit, G. *et al.* A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl Acad. Sci. USA* **99**, 5515–5520 (2002).
- Almasan, A. *et al.* Deficiency of retinoblastoma protein leads to inappropriate S-phase entry, activation of E2F-responsive genes, and apoptosis. *Proc. Natl Acad. Sci. USA* **92**, 5436–5440 (1995).
- Yin, Y. *et al.* Involvement of p85 in p53-dependent apoptotic response to oxidative stress. *Nature* **391**, 707–710 (1998).
- Maiyar, A. C., Huang, A. J., Phu, P. T., Cha, H. H. & Firestone, G. L. p53 stimulates promoter activity of the sgk. serum/glucocorticoid-inducible serine/threonine protein kinase gene in rodent mammary epithelial cells. *J. Biol. Chem.* **271**, 12414–12422 (1996).
- Minden, A. *et al.* c-Jun N-terminal phosphorylation correlates with activation of the JNK subgroup but not the ERK subgroup of mitogen-activated protein kinases. *Mol. Cell. Biol.* **14**, 6683–6688 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank A. J. Levine for advice on promoter studies and for providing critical reagents, and W. Gu for the pCMV-p53 Δ 270 plasmid. We are grateful to H. B. Lieberman for critical reading of the manuscript. We also thank T. A. Sato for help with the luciferase assay. This work was supported by a start up fund from the Columbia University (Y.Y.) and NIH grants (to Y.Y. and E.J.H.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.Y. (e-mail: yy151@columbia.edu).

The voltage dependence of NADPH oxidase reveals why phagocytes need proton channels

Thomas E. DeCoursey, Deri Morgan & Vladimir V. Cherny

Department of Molecular Biophysics and Physiology, Rush Presbyterian St Luke's Medical Center, 1750 West Harrison, Chicago, Illinois 60612, USA

The enzyme NADPH oxidase in phagocytes is important in the body's defence against microbes: it produces superoxide anions (O_2^- , precursors to bactericidal reactive oxygen species¹). Electrons move from intracellular NADPH, across a chain comprising FAD (flavin adenine dinucleotide) and two haems, to reduce extracellular O_2 to O_2^- . NADPH oxidase is electrogenic², generating electron current (I_e) that is measurable under voltage-clamp conditions^{3,4}. Here we report the complete current–voltage relationship of NADPH oxidase, the first such measurement of a plasma membrane electron transporter. We find that I_e is voltage-independent from -100 mV to >0 mV, but is steeply inhibited by further depolarization, and is abolished at about $+190$ mV. It was proposed that H^+ efflux² mediated by voltage-gated proton channels^{5,6} compensates I_e , because Zn^{2+} and Cd^{2+} inhibit both H^+ currents^{7–9} and O_2^- production¹⁰. Here we show that COS-7 cells transfected with four NADPH oxidase

components¹¹, but lacking H⁺ channels¹², produce O₂⁻ in the presence of Zn²⁺ concentrations that inhibit O₂⁻ production in neutrophils and eosinophils. Zn²⁺ does not inhibit NADPH oxidase directly, but through effects on H⁺ channels. H⁺ channels optimize NADPH oxidase function by preventing membrane depolarization to inhibitory voltages.

Because it is electrogenic, NADPH oxidase ought to be sensitive to membrane potential. However, this property has been demonstrated only over a limited voltage range^{3,13}. Furthermore, the hypothesis that voltage-gated proton channels compensate most of the charge generated by NADPH oxidase function^{2,5,6,14} has not been tested directly, and recently, K⁺ was proposed to have a similar role¹⁵. Although Zn²⁺ or Cd²⁺ can inhibit O₂⁻ production^{10,16}, they do so at higher concentrations than are required to inhibit voltage-gated proton currents^{5,9,17}, raising the possibility that these metals directly inhibit NADPH oxidase or affect other processes. Here we establish the importance of H⁺ currents, and show that the voltage dependence of NADPH oxidase optimizes its performance.

The current-voltage relationship of *I_e* generated by NADPH oxidase is shown in Fig. 1. Eosinophils in permeabilized-patch configuration^{4,17} were stimulated with the potent respiratory burst agonist, PMA (phorbol myristate acetate), which activates NADPH oxidase, resulting in inward *I_e* (ref. 17). We isolated *I_e* by its sensitivity to the NADPH oxidase inhibitor diphenylene iodonium chloride (DPI)¹⁸. Voltage-gated proton currents were inhibited by addition of 1.5–6.0 mM ZnCl₂ to the bathing solution. These high concentrations of ZnCl₂ (free Zn²⁺, 0.5–5.0 mM) did not reduce *I_e* noticeably. The mean reduction of *I_e* by 3 mM Zn²⁺ was 5.9 ± 6.5% (mean ± s.e.m., *n* = 6), ruling out direct inhibition of NADPH oxidase. Therefore, inhibition of O₂⁻ production by high Zn²⁺ concentrations in previous studies^{10,19} was not due to direct inhibition of NADPH oxidase. We compared currents elicited by voltage ramps soon after DPI addition (before inhibition was evident) and later (Fig. 1a, filled squares and open circles, respectively). Full inhibition of *I_e* by DPI takes 1–2 min (ref. 13). The net DPI-sensitive

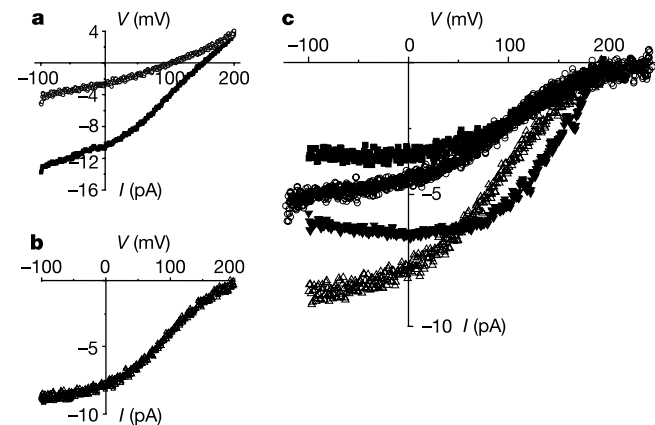


Figure 1 Current-voltage relationship for NADPH-oxidase-mediated electron current (*I_e*) in human eosinophils. **a**, Uncorrected averaged currents recorded in an eosinophil during voltage ramps early (filled squares) and late (open circles) after addition of DPI to a cell that was previously stimulated with 60 nM PMA. Voltage ramps were applied at 5-s intervals from -100 mV to +200 mV. The membrane was held at -60 mV after each ramp, and stepped to -100 mV for 400 ms before the start of the ramp. To inhibit H⁺ channels, 5 mM ZnCl₂ was added to all bath solutions. The average of three ramp currents recorded ~4 min after DPI was added was subtracted point-by-point from the average of four records obtained within 20 s after DPI addition. The difference (**b**) is DPI-sensitive *I_e*. Filled symbols indicate 'steady-state' net *I_e* measured at the end of the 400-ms prepulse to -100 mV and after 4.4 s at -60 mV. **c**, Net *I_e* in four eosinophils representative of the variability observed in *I_e*-V relationships in ten cells.

I_e is plotted in Fig. 1b. The similarity of *I_e* measured during ramps (open triangles) and at constant voltage (filled circles) confirms that *I_e* reached steady state at each voltage during the ramps. This result is expected because the turnover rate of NADPH oxidase is ~300 s⁻¹ (ref. 20).

I_e-V relationships measured in four cells (of ten studied) are plotted in Fig. 1c to illustrate the variability of the data. The expected inhibition of electron efflux by depolarization occurs only at large positive voltages, where the *I_e*-V relationship becomes steeply voltage dependent. Electron transport was abolished at +165 to +220 mV (+188 ± 19 mV; mean ± s.d., *n* = 10). The standard redox potential of the NADPH/NADP⁺ pair is -320 mV (ref. 21) and that of O₂/O₂⁻ is -160 mV (ref. 22). Thus, the nominal driving force for electron movement across the chain is +160 mV, although this pseudo-equilibrium potential ought to depend on concentration²³. Depolarization of the phagocyte membrane to the equilibrium potential of the electron transport pathway shuts down NADPH oxidase. The *I_e*-V relationship was surprisingly independent of voltage from -100 mV to an inflection point that occurred between 0 mV and +80 mV. In some cells there was weak voltage dependence at negative voltages (for example, open symbols in Fig. 1c), but strong rectification of the *I_e*-V relationship was observed consistently. Evidently, a voltage-independent process is rate limiting at voltages <0 mV, and a voltage-sensitive process becomes rate determining only during extreme depolarization. Electron transfer between the two intramembrane haem groups in gp91^{phox} (ref. 24) is a likely candidate for this voltage sensitive process.

If no compensatory mechanism existed, *I_e* in an eosinophil at 37 °C would depolarize the membrane by ~1 kV min⁻¹ during the respiratory burst, driving the membrane from its resting potential of -60 mV (ref. 25) to +190 mV in <20 ms, and O₂⁻ production would cease. For continuous function of NADPH oxidase, this charge movement must be balanced by an efficient mechanism that responds rapidly to depolarization. Voltage-gated proton channels are ideally suited to this task, because they are activated by cytosolic acidification and depolarization^{5,8,17,26}. Depolarization to +20 mV activates sufficient outward H⁺ current to compensate *I_e* completely in a PMA-stimulated eosinophil¹⁷. However, the only evidence supporting this role is the inhibition of O₂⁻ production by Cd²⁺

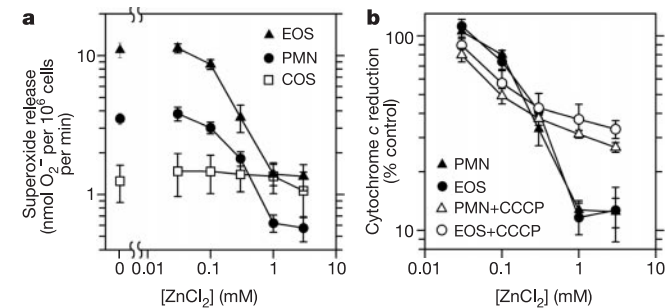


Figure 2 Factors affecting superoxide production. **a**, Sensitivity of the maximum rate of superoxide production to ZnCl₂ in human eosinophils (EOS, filled triangles), neutrophils (PMN, filled circles) and COS_{phox} cells (COS, open squares) (mean ± s.e.m. of 4, 9 or 6 experiments, respectively). Release of O₂⁻ at 37 °C was quantified by reduction of cytochrome c as described previously¹⁷, except that cells were suspended in Ringer's solution to avoid chelation of Zn²⁺. Inhibition was significant at [ZnCl₂] ≥ 0.3 mM (*P* < 0.01, Student's unpaired *t*-test). **b**, Restoration of O₂⁻ production (mean and s.e.m.) by 60 μM CCCP (open symbols) in human eosinophils (EOS, filled triangles, *n* = 3) or neutrophils (PMN, filled circles, *n* = 4) inhibited by Zn²⁺. For both cell types, O₂⁻ production at 1 or 3 mM ZnCl₂ was significantly greater in the presence of CCCP (*P* < 0.05). CCCP alone appeared to partially inhibit O₂⁻ release, perhaps by uncoupling mitochondria, although this was not statistically significant. The control data are a subset of the data in **a**. All values are normalized to those in the absence of Zn²⁺.

and Zn^{2+} (ref. 10) at concentrations larger than those needed to inhibit voltage-gated proton current^{9,16}. If Zn^{2+} inhibits O_2^- production by inhibiting H^+ currents, then it should not affect O_2^- production in cells that do not express voltage-gated proton channels. The only cells known to lack H^+ channels are COS-7 cells¹². Therefore, we studied O_2^- production in COS-7 cells transfected with the four main NADPH oxidase components, gp91^{phox}, p47^{phox}, p67^{phox} and p22^{phox}, which we call COS_{phox} cells^{11,12}. In Fig. 2 the concentration–response relationship for ZnCl_2 inhibition of the maximal rate of O_2^- production is compared in human eosinophils (filled triangles), neutrophils (filled circles) and COS_{phox} cells (open squares) stimulated with 60 nM PMA. O_2^- generation was measured by a standard cytochrome *c* reduction assay¹⁷. There was little effect of up to 3 mM ZnCl_2 on COS_{phox} cells, which lack H^+ channels. In contrast, 0.1–0.3 mM ZnCl_2 clearly reduced O_2^- production in human phagocytes, with nearly complete inhibition at 3 mM ZnCl_2 . Evidently, ZnCl_2 reduces O_2^- production by inhibiting voltage-gated proton channels, and cannot do so in cells that lack H^+ channels. The charge-compensating mechanism of COS_{phox} cells is not known, but it evidently is less sensitive to Zn^{2+} than are H^+ channels. Figure 2b illustrates that the inhibition of O_2^- production in phagocytes by Zn^{2+} was partially overcome by addition of the protonophore, CCCP (carbonyl cyanide *m*-chlorophenylhydrazone). In contrast, the K^+ ionophore, valinomycin, exacerbates the inhibition of O_2^- production by Cd^{2+} or Zn^{2+} (refs 6, 10). Together, these results indicate that proton efflux rather than flux of another ion compensates for charge separation by NADPH oxidase in phagocytes.

A small fraction of I_e might be compensated by K^+ , as was proposed recently¹⁵. However, we rule out significant involvement of K^+ current in compensating charge. First, although unstimulated eosinophils have inwardly rectifying K^+ channels, they have no outward K^+ conductance, and blocking the inward rectifier does not compromise O_2^- production²⁷. In PMA-stimulated eosinophils in which I_e attests to NADPH oxidase activity, we do not detect outward K^+ currents (data not shown). Furthermore, upon stimulation by PMA, the membrane potential of eosinophils depolarizes to the Nernst potential for H^+ over a wide range of pH gradients²⁸, demonstrating that the predominant conductance is proton selective. Thus, H^+ efflux compensates most of the charge separation by NADPH oxidase in active phagocytes. Although any K^+ efflux would contribute to charge compensation, the primary role of K^+ efflux is regulation of the phagosomal volume, ionic strength and pH (ref. 15). K^+ efflux, in contrast to H^+ efflux, causes osmotic and pH changes in the phagosome, which promote activity of proteolytic enzymes¹⁵. Similarly, complete depletion of phagosomal Cl^- by flux into cytosol would compensate less than 4% of the charge translocated by NADPH during the respiratory burst¹⁵. Proton flux is ideally suited to charge compensation, because it is osmotically neutral and pH neutral.

Voltage-gated proton channels are highly sensitive to divalent cations^{7–9}, with significant inhibition by 1 μM ZnCl_2 at pH ≥ 7 in eosinophils¹⁷. Paradoxically, much higher concentrations are required to inhibit O_2^- production. The I_e – V relationship reported here explains this apparent discrepancy. Zn^{2+} and Cd^{2+} do not ‘block’ H^+ channels by steric occlusion, but rather shift H^+ current activation to more positive voltages^{5,9,26}. Because I_e is voltage independent from the normal resting potential to beyond 0 mV (Fig. 1), shifting the threshold for H^+ current activation within this voltage range does not inhibit NADPH oxidase. Analogously, depolarization to ~ 0 mV with high K^+ concentration had no effect on O_2^- production by eosinophils²⁷, and only partially inhibited O_2^- production by neutrophils²⁹. Distinct inhibition of O_2^- production in neutrophils first occurred at 100–300 μM Zn^{2+} (Fig. 2), which would shift the threshold for activating H^+ channels at pH 7.4 by 80–90 mV (ref. 9); that is, from -20 mV in activated phagocytes⁴ to $+60$ to $+70$ mV. Consequently, activation of suffi-

cient H^+ efflux to compensate I_e would occur only at voltages within the range where I_e is inhibited directly by voltage. In conclusion, Zn^{2+} inhibits O_2^- production by shifting H^+ channel activation into or beyond the voltage-dependent region of the I_e – V relationship.

During the respiratory burst in neutrophils, the membrane depolarizes to $+58$ mV (ref. 30), which is close to the point at which depolarization begins to inhibit I_e . However, in spite of this depolarization of >100 mV from the resting potential, there is minimal ‘self-inhibition’, because I_e is practically voltage independent throughout this voltage range. The average reduction of I_e at $+58$ mV, relative to I_e at the ‘resting potential’ of -60 mV, was only $24 \pm 15\%$ (mean $\pm$ s.d., $n = 10$). The surprisingly large range over which NADPH oxidase is insensitive to membrane potential provides a safety factor that ensures optimal function of this enzyme unless it is confronted by drastic membrane depolarization. The depolarization that occurs during the respiratory burst is sufficient to activate compensatory H^+ efflux through H^+ channels without significantly inhibiting the NADPH oxidase. □

Methods

Eosinophil and neutrophil isolation

Venous blood was drawn from healthy adult volunteers under informed consent, as approved by our Institutional Review Board and in accordance with Federal regulations. Neutrophils were isolated by density-gradient centrifugation¹⁷, and were suspended in 10 mM HEPES-buffered HBSS (with Ca^{2+} and Mg^{2+}) at pH 7.4 for O_2^- measurements. Eosinophils were isolated from the neutrophils by negative selection using anti-CD16 immunomagnetic beads¹⁷. Patch-clamp studies were done on freshly isolated eosinophils, and on eosinophils incubated overnight at 37 °C in RPMI 1640 medium containing 25 mM HEPES and L-glutamine (Gibco), supplemented with 10% fetal bovine serum (Bio-Whittaker), 100 units ml^{-1} penicillin, 100 $\mu\text{g ml}^{-1}$ streptomycin (Sigma), and 1 ng ml^{-1} recombinant human GM-CSF (R & D Systems). Cells were provided by Larry L. Thomas.

COS_{phox} cells

COS-7 cells stably transfected with the four main subunits of NADPH oxidase—gp91^{phox}, p22^{phox}, p47^{phox} and p67^{phox} (COS_{phox} cells)—were developed by Price *et al.*¹¹ and were provided by M. Dinauer. COS_{phox} cells were maintained in suspension as described¹².

Electrophysiology

For permeabilized-patch recording, the pipette solutions contained 80 mM KCH_3SO_3 or 100 mM tetramethylammonium methanesulphonate, 50 mM NH_4^+ in the form of 25 mM $(\text{NH}_4)_2\text{SO}_4$, 2 mM MgCl_2 , 5 mM BES, 1 mM EGTA, titrated to pH 7.0, and $\sim 500 \mu\text{g ml}^{-1}$ solubilized amphotericin B (Sigma). The symmetrical 50 mM NH_4^+ gradient ‘clamped’ the intracellular pH to extracellular pH (ref. 4). The bath solution was identical to the 100 mM tetramethylammonium methanesulphonate pipette solution, but lacked amphotericin B. Studies were done at 20–25 °C. PMA and CCCP were obtained from Sigma. Other details of electrophysiological measurements are described elsewhere¹⁷.

Received 14 October 2002; accepted 25 February 2003; doi:10.1038/nature01523.

1. Babior, B. M. NADPH oxidase: an update. *Blood* **93**, 1464–1476 (1999).
2. Henderson, L. M., Chappell, J. B. & Jones, O. T. The superoxide-generating NADPH oxidase of human neutrophils is electrogenic and associated with an H^+ channel. *Biochem. J.* **246**, 325–329 (1987).
3. Schrenzel, J. *et al.* Electron currents generated by the human phagocyte NADPH oxidase. *Nature* **392**, 734–737 (1998).
4. DeCoursey, T. E., Cherny, V. V., Zhou, W. & Thomas, L. L. Simultaneous activation of NADPH oxidase-related proton and electron currents in human neutrophils. *Proc. Natl Acad. Sci. USA* **97**, 6885–6889 (2000).
5. DeCoursey, T. E. & Cherny, V. V. Potential, pH, and arachidonate gate hydrogen ion currents in human neutrophils. *Biophys. J.* **65**, 1590–1598 (1993).
6. Henderson, L. M., Chappell, J. B. & Jones, O. T. Internal pH changes associated with the activity of NADPH oxidase of human neutrophils. Further evidence for the presence of an H^+ conducting channel. *Biochem. J.* **251**, 563–567 (1988).
7. Mahaut-Smith, M. The effect of zinc on calcium and hydrogen ion currents in intact snail neurones. *J. Exp. Biol.* **145**, 455–469 (1989).
8. Thomas, R. C. & Meech, R. W. Hydrogen ion currents and intracellular pH in depolarized voltage-clamped snail neurones. *Nature* **299**, 826–828 (1982).
9. Cherny, V. V. & DeCoursey, T. E. pH-dependent inhibition of voltage-gated H^+ currents in rat alveolar epithelial cells by Zn^{2+} and other divalent cations. *J. Gen. Physiol.* **114**, 819–838 (1999).
10. Henderson, L. M., Chappell, J. B. & Jones, O. T. Superoxide generation by the electrogenic NADPH oxidase of human neutrophils is limited by the movement of a compensating charge. *Biochem. J.* **255**, 285–290 (1988).
11. Price, M. O. *et al.* Creation of a genetic system for analysis of the phagocyte respiratory burst: high-level reconstitution of the NADPH oxidase in a nonhematopoietic system. *Blood* **99**, 2653–2661 (2002).
12. Morgan, D., Cherny, V. V., Price, M. O., Dinauer, M. C. & DeCoursey, T. E. Absence of proton channels in COS-7 cells expressing functional NADPH oxidase components. *J. Gen. Physiol.* **119**, 571–580 (2002).

13. Cherny, V. V., Henderson, L. M., Xu, W., Thomas, L. L. & DeCoursey, T. E. Activation of NADPH oxidase-related proton and electron currents in human eosinophils by arachidonic acid. *J. Physiol. (Lond.)* **535**, 783–794 (2001).
14. Nanda, A. & Grinstein, S. Protein kinase C activates an H^+ (equivalent) conductance in the plasma membrane of human neutrophils. *Proc. Natl Acad. Sci. USA* **88**, 10816–10820 (1991).
15. Reeves, E. P. et al. Killing activity of neutrophils is mediated through activation of proteases by K^+ flux. *Nature* **416**, 291–297 (2002).
16. Kapus, A., Szász, K. & Ligeti, E. Phorbol 12-myristate 13-acetate activates an electrogenic H^+ -conducting pathway in the membrane of neutrophils. *Biochem. J.* **281**, 697–701 (1992).
17. DeCoursey, T. E., Cherny, V. V., DeCoursey, A. G., Xu, W. & Thomas, L. L. Interactions between NADPH oxidase-related proton and electron currents in human eosinophils. *J. Physiol. (Lond.)* **535**, 767–781 (2001).
18. Cross, A. R. & Jones, O. T. The effect of the inhibitor diphenylene iodonium on the superoxide-generating system of neutrophils. Specific labelling of a component polypeptide of the oxidase. *Biochem. J.* **237**, 111–116 (1986).
19. Lowenthal, A. & Levy, R. Essential requirement of cytosolic phospholipase A_2 for activation of the H^+ channel in phagocyte-like cells. *J. Biol. Chem.* **274**, 21603–21608 (1999).
20. Koshkin, V., Lotan, O. & Pick, E. Electron transfer in the superoxide-generating NADPH oxidase complex reconstituted *in vitro*. *Biochim. Biophys. Acta* **1319**, 139–146 (1997).
21. Burton, K. & Wilson, T. H. The free-energy changes for the reduction of diphosphopyridine nucleotide and the dehydrogenation of L-malate and L-glycerol 1-phosphate. *Biochem. J.* **54**, 86–94 (1953).
22. Wood, P. M. The redox potential of the system oxygen-superoxide. *FEBS Lett.* **44**, 22–24 (1974).
23. Läger, P. *Electrogenic Ion Pumps* (Sinauer Associates, Sunderland, Massachusetts, 1991).
24. Biberstine-Kinkade, K. J. et al. Heme-ligating histidines in flavocytochrome b_{558} : identification of specific histidines in gp91^{phox}. *J. Biol. Chem.* **276**, 31105–31112 (2001).
25. Gordienko, D. V. et al. Voltage-activated proton current in eosinophils from human blood. *J. Physiol. (Lond.)* **496**, 299–316 (1996).
26. Byerly, L., Meech, R. & Moody, W. Rapidly activating hydrogen ion currents in perfused neurones of the snail, *Lymnaea stagnalis*. *J. Physiol. (Lond.)* **351**, 199–216 (1984).
27. Tare, M. et al. Inwardly rectifying whole cell potassium current in human blood eosinophils. *J. Physiol. (Lond.)* **506**, 303–318 (1998).
28. Bánfi, B. et al. A novel H^+ conductance in eosinophils: unique characteristics and absence in chronic granulomatous disease. *J. Exp. Med.* **190**, 183–194 (1999).
29. Martin, M. A., Nauseef, W. M. & Clark, R. A. Depolarization blunts the oxidative burst of human neutrophils. Parallel effects of monoclonal antibodies, depolarizing buffers, and glycolytic inhibitors. *J. Immunol.* **140**, 3928–3935 (1988).
30. Jankowski, A. & Grinstein, S. A noninvasive fluorimetric procedure for measurement of membrane potential. Quantification of the NADPH oxidase-induced depolarization in activated neutrophils. *J. Biol. Chem.* **274**, 26098–26104 (1999).

Acknowledgements We thank A. R. Cross and L. L. Thomas for discussions, and T. Iastrebova and J. Murphy for technical assistance. This work was supported in part by the Heart, Lung and Blood Institute of the National Institutes of Health (T.E.D.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.E.D. (e-mail: tdecours@rush.edu).

Crystal structure of a transcription factor IIIB core interface ternary complex

Z. Sean Juo*, George A. Kassavetis†, Jimin Wang*, E. Peter Geiduschek† & Paul B. Sigler*‡

* Department of Molecular Biophysics & Biochemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA

† Center for Molecular Genetics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0634, USA

‡ Deceased

Transcription factor IIIB (TFIIIB), consisting of the TATA-binding protein (TBP), TFIIB-related factor (Brf1) and Bdp1, is a central component in basal and regulated transcription by RNA polymerase III^{1–4}. TFIIB recruits its polymerase to the promoter and subsequently has an essential role in the formation of the open initiation complex. The amino-terminal half of Brf1 shares a high degree of sequence similarity with the polymerase II general transcription factor TFIIB, but it is the carboxy-terminal

half of Brf1 that contributes most of its binding affinity with TBP^{5–8}. The principal anchoring region is located between residues 435 and 545 of yeast Brf1, comprising its homology domain II. The same region also provides the primary interface for assembling Bdp1 into the TFIIB complex⁹. We report here a 2.95 Å resolution crystal structure of the ternary complex containing Brf1 homology domain II, the conserved region of TBP and 19 base pairs of U6 promoter DNA. The structure reveals the core interface for assembly of TFIIB and demonstrates how the loosely packed Brf1 domain achieves remarkable binding specificity with the convex and lateral surfaces of TBP.

The crystal structure of the Brf1–TBP–DNA ternary complex is shown in Fig. 1. The revealed portion of Brf1 (residues 437–506) extends in parallel with helices H2 and H2' across the convex surface of TBP. After two nearly 90° turns, the remaining polypeptide positions itself along the lateral surface of the first structural repeat of TBP (Fig. 2a, b). This 'vine-on-a-tree' conformation is substantially different from previously reported polymerase (pol) II transcription ternary complexes involving TBP and DNA¹⁰. The structure also demonstrates that the convex surface of TBP is extensively engaged as the principal interface for binding. A total of 31 Brf1 residues in the crystal structure interact with 29 TBP residues through 31 hydrogen bonds and 114 van der Waals contacts. The large number of interactions is exceptional, approximately three to five times greater than for the comparable pol II ternary complex structures. As a result of all these contacts, 3,230 Å² of TBP surface area becomes inaccessible to solvent on Brf1 binding.

The conformation of Brf1 homology domain II revealed in the crystal structure is unusually extended, which does not fit the general idea of a structural domain. Given that this Brf1 segment, when unbound, is resistant to protease digestion (Z.S.J., unpublished results), it must adopt a stable structural fold in solution and undergo global unfolding on complex formation with TBP. This resembles the maintenance of a partially unfolded state of the GTPase-activating protein SptP by its cognate chaperone SicP, as the latter stabilizes the elongated helical conformation of the target protein¹¹. The revealed Brf1 domain does not involve DNA binding in the crystal structure. Notably, the unresolved segment (residues 507–596) contains a proposed cryptic DNA-binding domain¹². The H25 helix of Brf1 lies close to DNA and its trajectory points towards the start site of transcription, which is consistent with the DNA-binding potential of this segment.

The conformation of DNA-bound TBP remains similar to the related pol II ternary complex structures¹⁰, with an average root-mean-square (r.m.s.) deviation of 0.54 Å for Cα positions despite variations of sequence. In comparison with the apo-structure¹³, TBP undergoes noticeable conformational changes on DNA binding (averaged r.m.s. deviation of 1.19 Å for Cα positions). The two direct repeats move slightly closer to accommodate the widened narrow groove of the TATA element, and an approximately 3.3 Å displacement is observed at the first stirrup region of TBP when its second structural repeat is aligned. Owing to the pseudo-symmetrical nature of the yeast U6 TATA motif, its sequence for crystallization was modified to favour unidirectional binding of TBP¹⁴ (Fig. 2c). As observed previously, the TATA element is smoothly curved and doubly kinked at the termini. The minor groove width of the downstream B-form DNA is substantially narrower, which is a consequence of being predominantly (A + T)-rich. It is worth noting that the terminal A-tract segment interacts firmly with its non-crystallographic symmetry (NCS) counterpart (see Supplementary Information). Given that 20 base pairs of DNA stack between two complexes and maintain a stable conformation without any external support, the stiffness of A-tract sequence must contribute significantly to maintaining local structural stability in the crystal lattice¹⁴.

The Brf1 homology domain II is mainly helical. The existence of four helices (H21, H23, H24 and H25) is consistent with secondary

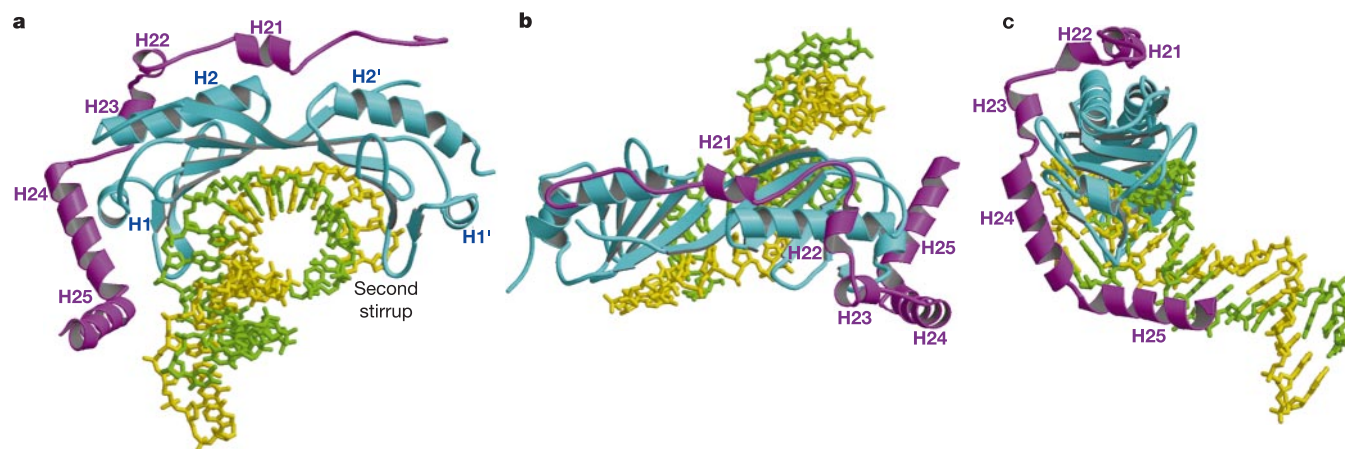


Figure 1 Ribbon style representations of the yeast Brf1-TBP-DNA ternary complex. TBP is coloured cyan and Brf1 is coloured magenta. DNA is shown in a ball-and-stick representation with the non-transcribed strand (the top strand in Fig. 2c) coloured green

and the transcribed (bottom) strand yellow. **a**, A 'front' view, with the first direct repeat of TBP to the left. **b**, A 'top' view, with the first direct repeat to the right. **c**, A 'side' view from the first direct repeat of TBP.

structure prediction¹⁵, with a precision of ± 1 residue. The only exception is helix H22 in the acidic region. The pattern of interaction between H22 and TBP closely resembles that of cyclic AMP-responsive activator CREB with co-activator CBP, in which the pKID domain of CREB undergoes a coil-to-helix folding transition on CBP binding¹⁶. If the formation of helix H22 is induced by interaction with TBP, this is probably due to four pairs of residues (R137/D464, K138/D467, R141/D466 and K145/D466; where TBP residues are the first in each pair) that establish a constraining framework at the circumference (Fig. 3a). A single residue, N491, is situated at the junction of helices H24 and H25 where the latter drifts away from the former by about 60°. The helical bend is mainly stabilized by two pairs of nonpolar contacts: W487-L495 within Brf1 and P96-F494 at the TBP-Brf1 interface (Fig. 3b). The hydrogen bond between K97 of TBP and D493 (or E497) of Brf1 may also contribute. It was reported that paired Brf1 mutations E506A/D508A result in temperature-sensitive phenotype¹⁷. It is probable that bending positions the C-terminal portion of Brf1 for optimal interaction with DNA, Bdp1, TFIIC or pol III subunits.

Yeast Brf1 homology domain II contains the principal anchoring

interface for TBP and Bdp1 (ref. 9). Two regions on the TBP surface are important for stable interaction with Brf1 (refs 18, 19): one is on the convex surface where the entire H2 helix of TBP is surrounded by Brf1 helices H21, H22, H23 and their intervening loops (Fig. 3a); the other is on the lateral surface, where helix H1 and the first stirrup of TBP engage Brf1 helices H24 and H25 (Fig. 3b). The large number of residues involved and their extensive contacts provide a structural basis for specific recognition between these two protein components.

Interactions on the convex surface of TBP are centred at helix H2, where the Brf1 binding sites cover a large portion of the first structural repeat of TBP. The principal cementing force is contributed by the acidic region of Brf1 (residues 457-469). This short stretch of peptide produces an exceptionally well-coordinated network of electrostatic interactions, within which an impressive 17 hydrogen bonds are packed densely. The central section of this acidic region is stabilized as helix H22 by TBP residues R137, K138, R141 and K145, and by two hydrophobic Brf1 residues (L462 and V465) that engage with the nonpolar surface of TBP (Fig. 3a). Moreover, TBP residues R137, K133 and K151 make six additional

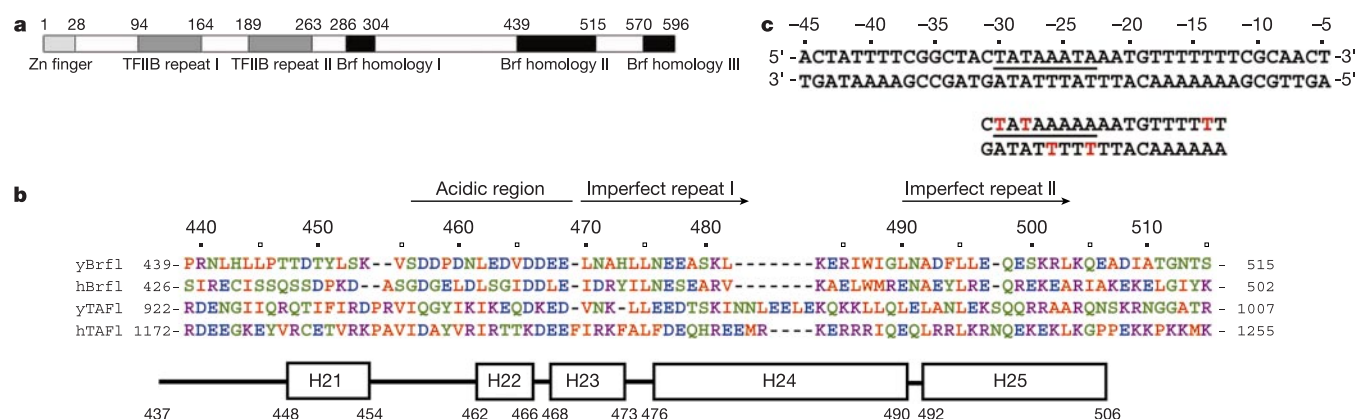


Figure 2 Brf1 and the yeast U6 promoter. **a**, The structural motifs of *Saccharomyces cerevisiae* Brf1. Sequence similarity of its N-terminal half with TFIIB and three conserved regions in the C-terminal half are indicated. **b**, Sequence alignment of yeast (y) and human (h) Brf1 homology domain II and the TAF1 HMG region. Residues are coloured on

the basis of side-chain properties; yeast Brf1 residue numbers are written along the top. **c**, The yeast U6 promoter sequence with the TATA element underlined. The 19-nucleotide duplex used for crystallization is shown below, with 5-bromo-dU substitutions for MAD phasing in red.

hydrogen bonds to properly maintain the trajectory of the loop region. Towards the N terminus, three polar contacts contributed by Brf1 T448 bridge helix H21 across the H2 and H2' helices of TBP. In a C-terminal position to helix H22, D467, E469 and N471 make several anchoring interactions that fix the position of helix H23. It is notable that this Brf1 segment, despite lacking a distinctive structural motif, achieves specific binding with TBP through a dense array of charge and shape complementarities at the protein interface.

In contrast to side-chain-specific interactions that are normally responsible for target recognition, eight of the ten hydrogen bonds on the lateral surface of TBP involve main-chain contacts. The only two TBP–Brf1 side-chain interactions, K83–E484 and K97–D493/E497, are of moderate strength, judging by the spatial separations. Therefore, the hydrogen bonds observed here contribute more towards the stability of global conformation than to the specificity of binding. Furthermore, hydrophobic interactions have an important role in reinforcing the local structure. The stack of van der Waals contacts between L87–W487 and Brf1 W487/L495 supplement the main-chain hydrogen bonds and fix the position of helix H24, whereas nonpolar interactions between P96 and F494 contribute towards stabilizing the helical bend between H24 and H25.

The amino acid sequence of Brf1 homology domain II is highly conserved. In addition to the strong resemblance of predicted secondary structures in this region between human and yeast Brf1, 28 of the 31 hydrogen bonds observed in the crystal structure can be retained at the human protein interface based on sequence conservation (Table 1). Two of the three missing hydrogen bonds are located on the lateral surface of TBP, where S480 and E484 of yeast Brf1 are both replaced by alanine in the human counterpart. The latter change converts an ionic interaction (with TBP K83) to a

nonpolar contact with TBP L87, and neither imposes any steric hindrance (Fig. 3b). The third missing hydrogen bond is located at helix H22 on the convex surface, where Brf1 residues E463 and D464 are replaced by serine and glycine in the human protein. The former allows a hydrogen bond with TBP R141, replacing the eliminated interaction with TBP R137 and Brf1 D464, and the latter facilitates the curving of the helical backbone. Thus, when human Brf1 is modelled into the structure of its yeast counterpart, the stabilization of helix H22 and the integrity of the TBP–Brf1 interface are preserved perfectly.

Interpretation of the structure–function relationship of this ternary complex benefits from a wealth of genetic and biochemical data. Applying regional codon randomization, one study identified an area on the convex surface of yeast TBP within which mutations of 19 residues result in pol III-specific temperature sensitivity that is partially suppressed by overexpressing Brf1 (ref. 20), suggesting potential roles of the corresponding TBP residues in Brf1 binding. The trace of involved residues is in marked accord with the crystal structure (Fig. 4a, panels (1) and (3)), and ten of these amino acids interact directly with Brf1 (Fig. 4b, lower panel, asterisks).

A systematic surface mutagenesis¹⁹ identified 15 human TBP residues whose mutations significantly reduce the binding of yeast Brf1 (Fig. 4a, panel (2); b, bottom panel). Nearly all of these residues are extensively engaged in Brf1 binding in the crystal structure. The most defective mutant, R137E (in yeast numbering), severely impairs the ability of TBP to interact with Brf1. Other mutations of TBP that substantially reduce Brf1 binding (25-fold or greater decrease) include L87E, H88E, K133E, K138Q and K145E. It has also been reported that K83E, L87E, H88E and K133E in yeast TBP are strongly defective for Brf1 interactions¹⁸. As revealed in the crystal structure, TBP residues K133, R137, K138 and K145 stabilize the H22 helix and/or properly anchor its surrounding loops (Fig. 3a); L87 involves numerous important van der Waals contacts that fix the position of helix H24 (Fig. 3b). TBP H88 does not engage in extensive side-chain interactions in the crystal structure; however, its human counterpart (R186) probably interacts with E477 of Brf1. Consequently, R186E of human TBP may repulse helix H24 from the lateral edge. All of these mutations directly correlate with TBP residues that define the global conformation of the interface with Brf1, and hence have a strong impact on Brf1 binding.

Mutagenesis of yeast Brf1 indicated that L462S, as well as two double mutations (D464A/D466A and K483A/E484A) decrease its affinity for TBP binding^{17,18} (Fig. 4c, asterisks). L462 interacts with five TBP residues (R137, A140, R141, Q144 and F152) and engages the highest number of nonpolar contacts in the crystal structure. D464 and D466 are largely responsible for the integrity of helix H22, whereas K483 and E484 are important for the position of helix H24 (Fig. 3). On the basis of the crystal structure, it is no surprise that these mutations are detrimental to TBP binding. Thus, we have observed excellent correlations between the structure and previously reported molecular genetic analyses.

It is important to note that this crystal structure also has potential implications for the TBP–TAF1 interaction. TAF1 (ref. 21) is a major scaffold component of the TFIID complex and serves, in part, as the principal intermediary between TBP and the remaining TAF subunits. A recent surface mutation analysis of TBP demonstrated that one of the TBP-anchoring interfaces of human TAF1 is situated in its high mobility group (HMG) homology region (residues 1172–1344)²². Nine of the twenty TBP residues whose mutations decrease full-length human TAF1 binding (by 40% or more) also affect interactions with this HMG segment²², and all nine residues (R107, K133, L134, R137, K138, R141, K145, K151 and F152, in yeast numbering) are in contact with yeast Brf1 in the crystal structure (Fig. 4a, panels (1) and (4)).

It has been reported that the human Brf1 homology domain II displays sequence similarity to chicken HMG-2 (ref. 6) and shares

Table 1 Hydrogen bonds and electrostatic interactions at the TBP–Brf1 interface

TBP atom ↔ Brf1 atom	hBrf1	Conservation yTAF1	hTAF1
Lateral edge			
K83 NZ ↔ E484 OE2	X	0	0
A86 O ↔ K483 NZ	0	0	0
L87 O ↔ S480 OG	X	0	X
H88 O ↔ L475 N	0	0	0
A89 O ↔ K483 NZ	0	0	0
R90 N ↔ H473 O	0	0	0
R90 NH1 ↔ A472 O	0	0	0
A92 O ↔ K483 NZ	0	0	0
Y94 O ↔ W487 NE1	0	X	0
K97 NZ ↔ D493 OD1	0	X	X
Convex surface			
R107 NE ↔ E469 OE2	0	0	0
R107 NH1 ↔ E469 OE2	0	0	0
R107 NH1 ↔ E469 OE1	0	0	0
E129 OE2 ↔ T448 OG1	0	0	0
K133 NZ ↔ D457 OD1	0	0	0
R137 NE ↔ S456 O	0	0	0
R137 NH1 ↔ D464 OD1	X	0	0
R137 NH2 ↔ D458 N	0	0	0
R137 NH2 ↔ D458 O	0	0	0
R137 NH2 ↔ S456 O	0	0	0
K138 NZ ↔ V465 O	0	0	0
K138 NZ ↔ D467 OD1	0	X	0
R141 NE ↔ D466 OD2	0	0	X
R141 NH2 ↔ D466 OD1	0	0	X
R141 NH2 ↔ D466 OD2	0	0	X
K145 NZ ↔ N471 OD1	0	0	X
K145 NZ ↔ D466 OD1	0	0	X
K151 NZ ↔ D460 OD1	0	0	0
F152 N ↔ P459 O	0	0	0
R220 N ↔ T448 OG1	0	0	0
E221 OE2 ↔ T448 N	0	0	0

A zero indicates that this specific TBP–Brf1 interaction observed in the crystal structure can be preserved at the specified TBP interface; X indicates that such an interaction cannot be transposed. The hydrogen bond with D460 can be maintained by interaction with neighbouring targets (E446 of hBrf1, Y944 of yTAF1 or Y1194 of hTAF1; see alignment in Fig. 2b). The double-headed arrows indicate an interaction. All interactions are shown in Fig. 3. (See Supplementary Information for hydrogen bond distances.)

sequence similarity with the HMG region of human TAF1 (ref. 22) (Fig. 2b). Notably, 28 of the 31 electrostatic interactions between TBP and Brf1 in the crystal structure can be transposed directly to the corresponding HMG segment of yeast TAF1 (Table 1). The transposition to human TAF1 is also reasonably successful (24 out of 31 interactions), but the important interactions with TBP R141 and K145, defined by mutation analysis²², cannot be retained. This should not seriously affect the overall binding affinity of the TAF1 HMG segment because most of the hydrogen bonds can still be preserved. Therefore, we anticipate largely conserved, but slightly different structural arrangements on the convex surface of the human TBP–TAF1 and TBP–Brf1 interfaces.

As the HMG region of TAF1 is likely to interact with TBP in a manner resembling the Brf1 homology domain II, a model of binding provides additional insights regarding how pol III transcription factors might initiate their assembly on TBP (Fig. 3c, d). The

putative TAF1 fragment, represented by the corresponding Brf1 domain, can largely accommodate the presence of TFIIA and TFIIB. However, there is local steric interference between residues 72–75 of the small subunit of TFIIA and residues 486–491 of the modelled fragment due to overlapping interactions with the first stirrup of TBP. Of note, the buckling force at the C-terminal half of helix H24 is largely released owing to the replacement of W487, D493, F494 and E497 (Fig. 3b). Consequently, this allows moderate movement for H24 to accommodate the incoming TFIIA, which is consistent with a previous prediction²². It is also possible that a slight rotation of TFIIA can avoid this apparent steric clash because the first stirrup of TBP is not its principal anchoring surface. This would potentially reposition TFIIA for closer engagement with upstream DNA.

We report here the crystal structure of a pol III-specific general transcription assembly. The structure clearly demonstrates that TBP

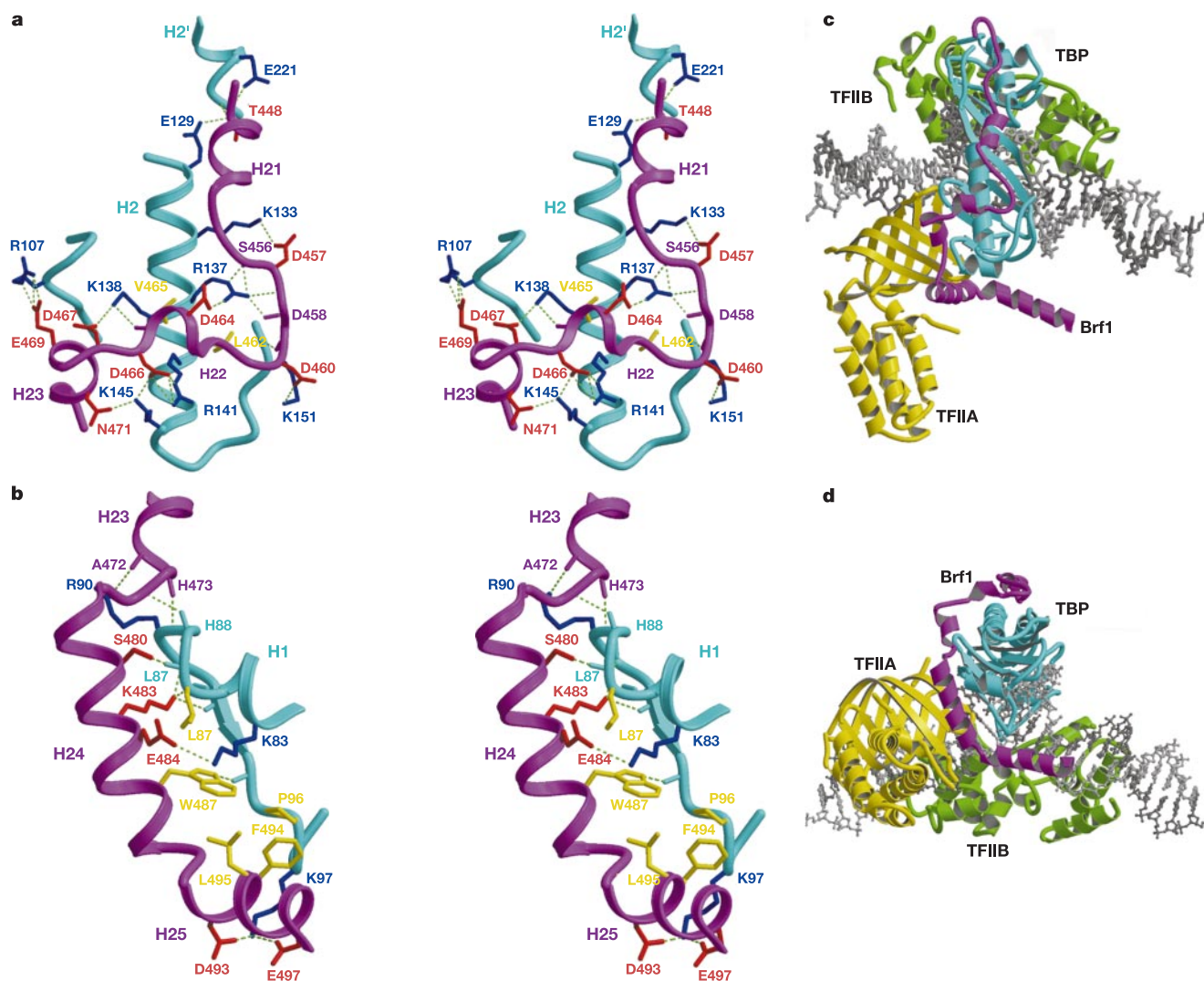


Figure 3 Stereo illustrations of the TBP–Brf1 interactions (left panel). The main chain of TBP and the main-chain interactions are coloured cyan whereas the side chains are blue; the main chain of Brf1 is coloured magenta and the side chains are red. Important hydrophobic residues are coloured yellow. Hydrogen bonds are indicated with green dotted lines. For clarity, only residues involved in crucial contacts are shown. **a**, A top view from the convex surface of TBP showing the extent of electrostatic interactions. **b**, A side view showing extensive main-chain hydrogen bonds and nonpolar interactions on the lateral surface of TBP. K97 of TBP, situated at the tip of the first stirrup, interacts with

either D493 or E497. The right panels (**c**, **d**) show a model indicating how the HMG region of TAF1 might orient itself on TBP with respect to other pol II general transcription factors. Proteins are shown in ribbon style, with TBP cyan, TFIIA yellow, TFIIB green and the putative TAF1 fragment (represented by the Brf1 domain) magenta. DNA strands (downstream to the right) are in a ball-and-stick representation with the top strand in light grey and the bottom strand in dark grey. The orientations of presentation are similar between **a** and **c**, and between **b** and **d**.

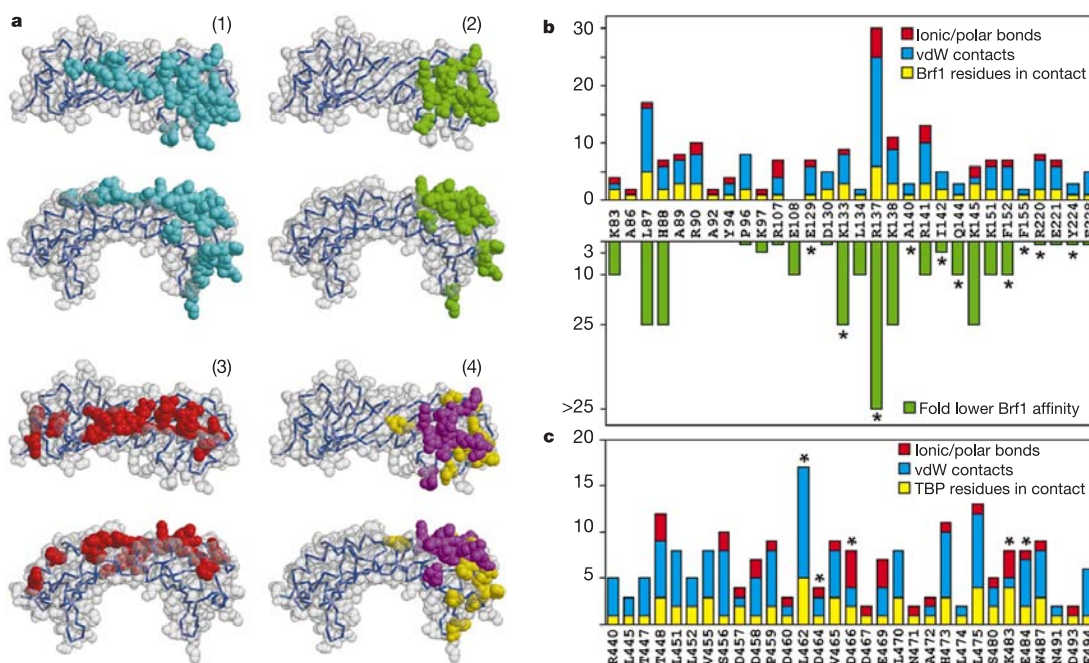


Figure 4 Cross-validation with molecular genetic analysis. **a**, A semi-transparent space-filling model with blue backbone tracing, showing yeast TBP in two orientations: a top view in the same orientation of Fig. 1b and a back view from the direction opposite to Fig. 1a. Panel (1): all 29 TBP residues interacting with yeast Brf1 in the crystal structure are coloured cyan; panel (2): the 15 TBP residues whose mutations impair yeast Brf1 binding (<33%)¹⁹ are coloured green; panel (3): the 19 TBP residues whose mutations result in pol III-specific temperature sensitive phenotype²⁰ are coloured red; panel (4): the 9 TBP residues whose mutations reduce human TAF1 binding (<60% for full-length protein and for the HMG region)²² are coloured magenta; another 11 TBP residues whose mutations

recognizes pol II and pol III general transcription factors through different regions of its surface, despite their sequence similarity. The crystal structure also demonstrates extensive and highly coordinated interactions with a region of the convex surface of TBP that is known to be important for transcriptional activation. The new structural information provides an excellent cross-validation with a wealth of molecular genetic data reported in the past decade, and also presents supporting evidence for a potential conservation of TBP–TAF interactions between TFIID and TFIIB. □

Methods

Protein purification and crystallization

Brf1 (residues 437–596) was engineered to contain cleavable poly-histidine and glutathione *S*-transferase affinity tags (at the N and C termini, respectively) to facilitate purification. The fusion protein was produced in *Escherichia coli* using the pET expression system. Cells were grown at 30 °C and the temperature was reduced to 20 °C before induction with isopropyl- β -D-thiogalactoside (final concentration 0.1 mM) for 10 h. The target protein was purified under native conditions through Ni-NTA agarose (Qiagen) and glutathione Sepharose (Pharmacia). On tag cleavage, the final product was obtained by chromatography on Q Sepharose HP (Pharmacia). Expression and purification of the yeast TBP core domain (residues 61–240) was reported previously²³. The yeast Brf1–TBP–DNA ternary complex was concentrated to 0.2 mM in a buffer containing 10 mM MES (pH 6.5), 100 mM NaCl, 2 mM 2-mercaptoethanol and 10% (v/v) glycerol. The crystals were grown in a period of 3–5 weeks using the hanging-drop vapour diffusion method at 4 °C by mixing 1:1 volume ratio of the complex with the following cryo-ready precipitant: 50 mM MES (pH 6.5), 100 mM NaCl or LiCl, 10 mM MgCl₂, 2 mM 2-mercaptoethanol, 25% (v/v) glycerol and 10% (v/v) polyethylene glycol 2000 mono-methyl ether. The crystals belong to orthorhombic space group *P*2₁2₁2, with unit cell dimensions of *a* = 93.6 Å, *b* = 52.6 Å, *c* = 256.6 Å.

Structure determination and refinement

Diffraction data were processed using Denzo and Scalepak in the HKL2000 program suite²⁴. A multiple wavelength anomalous dispersion (MAD) experiment was performed on a derivative crystal in which five thymine bases were replaced by 5-bromouracil.

only affect full-length human TAF1 binding (but not the HMG region) are coloured yellow. **b**, Quantitative analysis of the 29 TBP residues interacting with Brf1 in the crystal structure. Above the sequence: the numbers of Brf1 residues engaged, van der Waals (vdW) contacts and hydrogen bonds are coloured yellow, cyan and red, respectively; below the sequence: the effect of each mutation on Brf1 binding¹⁹ is indicated in green. (No effect is shown as onefold; the absence of a green bar indicates no data.) The pol III-specific temperature-sensitive mutations²⁰ are labelled with asterisks. **c**, Quantitative analysis of the 31 Brf1 residues interacting with TBP in the crystal structure. The Brf1 residues whose mutations affect TBP binding are labelled with asterisks^{17,18}.

Experimental phase was calculated with the program SHARP²⁵, and additional density modification procedures such as 'data sharpening' and NCS averaging were applied, rendering the maps interpretable. Electron densities of the Brf1 homology domain II along with the TBP–DNA portion of the ternary complex were clear in the experimental map and were continuous throughout the 2|*F*_o| – |*F*_c| difference maps; however, the C-terminal portion of the Brf1 fragment (beyond residue 507) was missing in the density. Structural refinement was carried out using complete data with no resolution or sigma cutoffs. Simulated annealing, positional and *B*-factor refinements were performed using the program CNS²⁶. NCS-related movement was adjusted with TLS refinement²⁷ through the program REFMAC²⁸. The final statistics (*R*_{work} = 27.7%, *R*_{free} = 30.8%, r.m.s. deviation bond length = 0.010 Å, r.m.s. deviation bond angle = 1.441°) are reasonable, given the resolution limit and the anisotropic diffraction of the native data. PROCHECK²⁹ analysis of the refined model indicated that only four residues (0.5%) are classified as unfavourable in the Ramachandran plot. The main-chain and side-chain geometry statistics are either better than or compatible with other structures of similar resolution, with an overall *G*-factor of –0.2 compared with a typical value of –0.8 (±0.3). See Supplementary Information for complete crystallographic statistics.

Received 27 December 2002; accepted 11 March 2003; doi:10.1038/nature01534.
Published online 26 March 2003.

- White, R. J. *RNA Polymerase III Transcription* (Springer/Landes Bioscience, New York/Georgetown, 1998).
- Geiduschek, E. P. & Kassavetis, G. A. The RNA polymerase III transcription apparatus. *J. Mol. Biol.* **310**, 1–26 (2001).
- Schramm, L. & Hernandez, N. Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16**, 2593–2620 (2002).
- Willis, I. M. A universal nomenclature for subunits of the RNA polymerase III transcription initiation factor TFIIB. *Genes Dev.* **16**, 1337–1338 (2002).
- Khoo, B., Brophy, B. & Jackson, S. P. Conserved functional domains of the RNA polymerase III general transcription factor BRF. *Genes Dev.* **8**, 2879–2890 (1994).
- Wang, Z. & Roeder, R. G. Structure and function of a human transcription factor TFIIB subunit that is evolutionarily conserved and contains both TFIIB- and high-mobility-group protein 2-related domains. *Proc. Natl Acad. Sci. USA* **92**, 7026–7030 (1995).
- Chaussivert, N., Conesa, C., Shaaban, S. & Sentenac, A. Complex interactions between yeast TFIIB and TFIIC. *J. Biol. Chem.* **270**, 15353–15358 (1995).
- Ruth, J. *et al.* A suppressor of mutations in the class III transcription system encodes a component of yeast TFIIB. *EMBO J.* **15**, 1941–1949 (1996).
- Kassavetis, G. A., Kumar, A., Ramirez, E. & Geiduschek, E. P. Functional and structural organization of

- Brf, the TFIIB-related component of the RNA polymerase III transcription initiation complex. *Mol. Cell Biol.* **18**, 5587–5599 (1998).
10. Patikoglou, G. & Burley, S. K. Eukaryotic transcription factor-DNA complexes. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 289–325 (1997).
11. Stebbins, C. E. & Galan, J. E. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **414**, 77–81 (2001).
12. Huet, J., Conesa, C., Carles, C. & Sentenac, A. A cryptic DNA binding domain at the COOH terminus of TFIIB70 affects formation, stability, and function of preinitiation complexes. *J. Biol. Chem.* **272**, 18341–18349 (1997).
13. Chasman, D. I., Flaherty, K. M., Sharp, P. A. & Kornberg, R. D. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proc. Natl Acad. Sci. USA* **90**, 8174–8178 (1993).
14. Juo, Z. S. *et al.* How proteins recognize the TATA box. *J. Mol. Biol.* **261**, 239–254 (1996).
15. Rost, B. PHD: predicting 1D protein structure by profile based neural networks. *Methods Enzymol.* **266**, 525–539 (1996).
16. Radhakrishnan, I. *et al.* Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* **91**, 741–752 (1997).
17. Andrau, J. C., Sentenac, A. & Werner, M. Mutagenesis of yeast TFIIB70 reveals C-terminal residues critical for interaction with TBP and C34. *J. Mol. Biol.* **288**, 511–520 (1999).
18. Colbert, T., Lee, S., Schimmack, G. & Hahn, S. Architecture of protein and DNA contacts within the TFIIB-DNA complex. *Mol. Cell Biol.* **18**, 1682–1691 (1998).
19. Shen, Y., Kassavetis, G. A., Bryant, G. O. & Berk, A. J. Polymerase (Pol) III TATA box-binding protein (TBP)-associated factor Brf binds to a surface on TBP also required for activated Pol II transcription. *Mol. Cell Biol.* **18**, 1692–1700 (1998).
20. Cormack, B. P. & Struhl, K. Regional codon randomization: defining a TATA-binding protein surface required for RNA polymerase III transcription. *Science* **262**, 244–248 (1993).
21. Tora, L. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* **16**, 673–675 (2002).
22. Martel, L. S., Brown, H. J. & Berk, A. J. Evidence that TAF-TATA box-binding protein interactions are required for activated transcription in mammalian cells. *Mol. Cell Biol.* **22**, 2788–2798 (2002).
23. Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127–151 (1996).
24. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
25. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
26. Brunger, A. T. *et al.* Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
27. Winn, M., Isupov, M. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements Z.S.J. would like to thank T. Steitz for financial support. The authors would like to thank A. Berk, J. Cabral, S. Harrison, N. Hernandez, F. Li, K. Struhl and Y. Xiong for discussions; G. Olack for assistance in mass spectrum analysis; M. Wilson for help with data collection; P. Mann for assistance with protein preparation; Yale CSB core staff for advice on software-related issues; W. Henderickson, C. Ogata (NSLS beamline X4A), M. Becker, L. Berman, R. Sweet (NSLS beamline X25), A. Joachimiak and SBC staff (ANL beamline 19 ID) for help with synchrotron data acquisition; and Z. Otwinowski, W. Minor, I. Minor, C. Vonnheim and G. Murshudov for advice on programs. This work was supported in part by the Howard Hughes Medical Institute and by the National Institutes of Health (E.P.G., P.B.S. and T. Steitz).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Z.S.J. (e-mail: juo@csb.yale.edu). Atomic coordinates have been deposited in the Brookhaven Protein Data Bank under accession number 1NGM.

Programs for action

New software packages tackle graphs, genomes, stability.

Grapher 4

Golden Software www.goldensoftware.com

Grapher 4 gains an extra dimension

The upgrade to the Grapher graphing package gives it the ability to create 3D graphs and to embed an Excel window directly within Grapher. These features enable users to create publication-quality graphs with the worksheet flexibility of the spreadsheet program. Users can convert their data into one of twenty-two fully-customizable graph types. Other new features are date/time support, allowing dates and times to be used as variables, and the ability to break an axis. It imports and exports data files in all popular formats, and its functionality can be integrated with any ActiveX Automation-compatible software. Visual Basic, C++ and Perl scripts can be used to call Grapher to automate repetitive functions.

Image-Pro Family Software Suite

Media Cybernetics www.mediacy.com

Image analysis solutions with a software suite

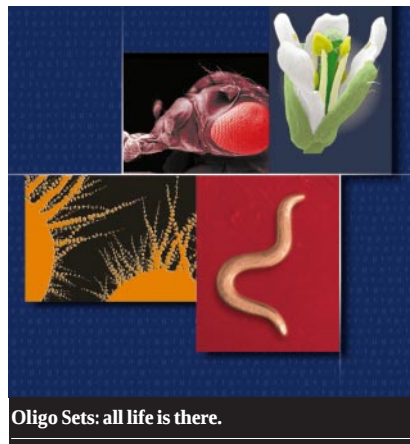
Media Cybernetics, a provider of scientific image analysis software, has expanded on its products to offer the Image-Pro Family software suite with a three-step imaging approach, allowing users to upgrade as their needs and imaging skills grow. Image-Pro Express is for basic imaging or for image capture stations, offering numerous capture and enhancement options. Added measurement and analysis capabilities can be found with the Image-Pro Discovery. And Image-Pro Plus includes all of the functionality of Image-Pro Discovery along with added analysis tools and the ability to write customized macros and plug-ins. Image-Pro products are also available in integrated camera and hardware kits.

OmniSEC

Viscotek www.viscotek.com

Software for chromatography automation

OmniSEC is a fully 21CFR part 11 compliant GPC/SEC chromatography software package providing an integrated software solution that helps maximize the information extraction and efficiency of polymer characterization experiments. It can be used to fully automate the Viscotek GPCmax, a pump, autosampler and degasser module designed for polymer characterization applications. OmniSEC could be used in laboratories where it is of primary importance to provide integrated system control and data acquisition from the same software with traceable data. All system parameters need only be entered once, enabling secure and traceable data entry.



Oligo Sets: all life is there.

Array-Ready Oligo Sets

Qiagen www.operon.com

Four new genome sets

Arabidopsis, *Candida albicans*, *Caenorhabditis elegans* and *Drosophila melanogaster* genomes have been added to Qiagen's line of Oligo Sets. Each 70-mer probe is designed for maximal specificity for its target gene, while minimizing cross-hybridization to other genes. Each probe is melting-temperature normalized and con-

tains a 5' amino linker, giving the flexibility to select covalent or noncovalent slide chemistry. Array-Ready Oligo Sets are offered in a 384-well plate format at 600 or 1,200 pmol.

Stability Pre-Scheduling Module

Analytical Information Systems

www.ais-lims.com

LIMS software for stability

Now a stability trial can be managed entirely by laboratory information management systems. The Stability Pre-Scheduling module allows samples to be pre-registered onto the LIMS for planned release at predetermined times. The Analytical Information Systems LIMS organizes the work protocol by informing lab staff of what resources are required for the stability study, what needs to be done and when to do it. New tools allow authorized users to modify screens and data structures. If the stability study starts to suffer data drift, the user is immediately informed, improving the quality of the stability trial and removing human error in spotting some mistakes.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Senior European Sales Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin

To send materials use London

address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Young at heart

A few weeks ago, the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, resembled a scientific version of a football training camp. A hundred or so prospective graduate students displayed an air of excitement as they huddled around possible mentors and congregated at the institute's double-helix fountain. But there was a sense of anxiety, too, because by the end of the week, only half of them would be offered a studentship slot for the autumn — the rest would go home empty-handed, the equivalent of not making the team.

The visitors had already been screened down from about 420 applicants — a sizeable response that was not lost on the European Research Council (ERC). Earlier last month, the ERC met to discuss how to reshape European academic research to attract students and be competitive with labs in the United States. It cited EMBL as a model for the initiative.

So what is the secret of EMBL's success? Iain Mattaj, the institute's scientific coordinator, says that its turnover system plays a large role. Scientific group leaders tend to be young — usually having just finished their first postdoc — and work under five-year contracts, which are renewable for up to four more years. This means that the lab has a constant influx of new ideas and keeps scientific stagnation at bay — qualities that attract graduate students. "The pluses of having the turnover system are enormous," says Mattaj. "It brings a tremendous vitality."

But, he admits, there are downsides, too. Perhaps the biggest is that the turnover and youthfulness are far less attractive to prospective postdocs, who want mentors who are more established. Even so, the ability to attract energetic students — and the palpable energy EMBL emits during 'draft day' — is something other European research institutes should consider emulating.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

Changing attitudes in the world
of drug discovery p542

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

The biotechnology sector's craving for commercial expertise is fuelling a cultural exchange with the drug industry. Paul Smaglik and Simon Frantz investigate.



Josef Hermanns stresses that communication is the key to success, no matter how big the business is.

Pharmaceutical companies are lumbering, bureaucratic dinosaurs; biotechnology firms, on the other hand, are nimble, innovative entrepreneurial mammals. These are the common stereotypes scientists tend to hold about the two arms of drug discovery. But they are distinctions that are dissolving, says Mark Wuonola, director of infection strategy and management at AstraZeneca in Waltham, Massachusetts.

Wuonola should know. He is one of those rare people who has gone from pharmaceuticals to biotech and back again. Many scientists make the switch from the drug industry to the biotech sector, but far fewer make the move the other way.

Wuonola's first job, at DuPont Merck in Wilmington, Delaware, saw him progress from chemistry to strategic planning to patents. His role at biotech firm Scriptgen in Waltham, Massachusetts, saw him embrace all of these roles — and more — almost immediately. Now back in the drug sector, Wuonola is keen to see a stronger shift in corporate values. Biotech firms could stand to learn from the pharmaceutical sector's more experienced management and history of getting drugs through the regulatory process. Drug companies could learn from the biotech sector's ability to move quickly and think creatively.

Such cross-fertilization has already begun. For example, AstraZeneca's Boston site resembles a biotech in that it has a relatively small number of scientists working on a few targeted areas. Meanwhile, smaller firms are turning to the pharmaceutical sector for management expertise and people with track records of ushering compounds through trials and into the marketplace. At London-based biotech firm Trigen, for example, 6 of its 23 staff hail from big drug companies.

Having people who have pushed compounds through the pharmaceutical pipeline helps to introduce other employees to fresh aspects of the drug-discovery process. That is particularly helpful if the bulk of employees come from an academic background, as is the case at Cellzome, a drug-discovery firm based in Heidelberg, Germany.

"Everyone in a company like this is on a steep learning curve," says Giulio Superti-Furga, Cellzome's senior vice-president for biology. Spun off from the European Molecular Biology Laboratory in 2000, Cellzome now employs 95 people and, like many others in the biotech sector, is turning to the

industrialized world of pharmaceuticals in search of the commercial expertise it needs for success. Last year, David Brown joined the firm as president and chief executive after 25 years in the pharmaceutical sector, including spells at Roche and GlaxoWellcome.

GROWTH ISSUES

Differences in communication style are probably one of the major contrasts between the two sectors. Franz Hefti, who recently joined biotech start-up Rinat Neuroscience in Palo Alto, California, as executive vice-president of development, sensed a difference immediately. Before joining Rinat, Hefti was senior vice-president for neuroscience research at drug firm Merck. Compared with Merck, "there is more openness and a higher degree of cooperation", Hefti says, "which makes certain processes more dynamic." As a result, Hefti finds that it is now possible to solve problems through simple face-to-face conversations.

One of the obvious factors feeding this dynamic is size — Rinat employs around 25 people. "The first thing I noticed when I started at Rinat was that everything was much easier to survey," says Hefti. "Within a day I got to know everyone there and their skills, a situation that is unthinkable in a large pharmaceutical company." The downside is that this means there is a smaller pool of expertise to draw on. "You don't have the large group of world-leading experts that you find in larger companies," Hefti says.

Fuelled by fresh expertise and with products on the



Giulio Superti-Furga: expertise from the drug sector can boost biotech.



Scott Sneddon found the switch from big industry to the smaller world of biotechnology gave him more diverse roles.

market, some biotech firms are finding that success — and the attendant growth — brings its own challenges. Josef Hermanns has experienced the rollercoaster ride. For the past six years, Hermanns has worked for a 10-person start-up, a 100-person bioinformatics company, a 500-person drug-development firm, and a 380-person bioinformatics concern. All four companies bore the same name: LION Bioscience.

For Hermanns, who joined the Heidelberg-based company in 1997 and is now project director for scientific content, each entity might as well have carried a different name. The evolution between them has been “drastic,” he says.

“When you start with ten people, processes are intuitive,” Hermanns says. “This changes when you move to 100.” At that point, informal dialogue is not as effective as the vibrant give-and-take when the company is embryonic. “You enter a stage where you start to have communication problems,” he acknowledges.

If communications are problematic as biotech firms grow, the difficulties are multiplied within the globalized

players in the drug industry. That is why Merck tries to keep its research groups down to a manageable size, focused on specific themes, says Gerry Dawson, senior director of behavioural neuroscience at Merck's site in Terlings Park, UK. “You don't feel like you're in a big company,” he says.

In fact, this focused mentality can leave employees in large companies feeling refreshingly free of a ‘corporate mentality’. Basic research groups can function in a way that is much more removed from the business than is possible in a small company, says Scott Sneddon, who used to work at drug firm Pfizer in Connecticut as a computational chemist. But Sneddon feels that this can leave employees chained to the bench — a factor in his move to become senior scientific director at biotech firm Genzyme in Cambridge, Massachusetts (see ‘A change of pace’, left).

David Alker, discovery recruitment and academic liaison manager for Pfizer in Sandwich, UK, agrees — to a point. Compared with the pharmaceutical sector, researchers in biotech firms tend to move more quickly into areas beyond the bench, such as project management or marketing, he says. But the career path within the drug industry allows employees to hone a particular skill set, he explains.

One of the biggest advantages of working for a larger company is the opportunity for formal training, rather than on-the-job learning, says Dawson. That is how Merck is meeting its need for people who can test compounds in animals. “We run our own *in vivo* training scheme,” Dawson says.

Daniel Burns, vice-president of discovery genetics at GlaxoSmithKline in Research Triangle Park, North Carolina, feels that people considering the switch from big industry to smaller firms should weigh up the options carefully. “In large pharmaceutical firms there are several projects on the go, and we say that this isn't a spectator sport; you've got to be in the game,” he says. In small companies, there are fewer projects, a narrower focus and a greater sense of familiarity with co-workers. Some people prefer the high level of interaction and clarity of purpose this arrangement offers, Burns notes.

So as biotech firms and pharmaceutical companies grow to resemble each other more and more, prospective employees will need to seek out the remaining differences in order to find the location that suits them the best. ■

Paul Smaglik is editor of *NatureJobs*; **Simon Frantz** is associate editor, news, at *Nature Reviews Drug Discovery*.

A change of pace

Scott Sneddon switched from drugs to biotechnology because he wanted more interaction and a chance to assume different roles. After five years at Pfizer in Connecticut as a computational chemist, he felt he had glimpsed the whole drug-development chain, but was relegated to working within his link.

“I looked within the organization and saw the contributions that people in high-throughput screening and screening in biological assays were

making,” he says. “Those people were making important contributions in getting compounds into clinical trials and I wanted to be part of that process.”

Sneddon joined Genzyme in Cambridge, Massachusetts, as senior scientific director in 1995. One of the biggest advantages of the move, he says, is that he now has the chance to wear many different hats — from commercial to scientific. He also enjoys following one compound through the clinical pipeline.

P.S.